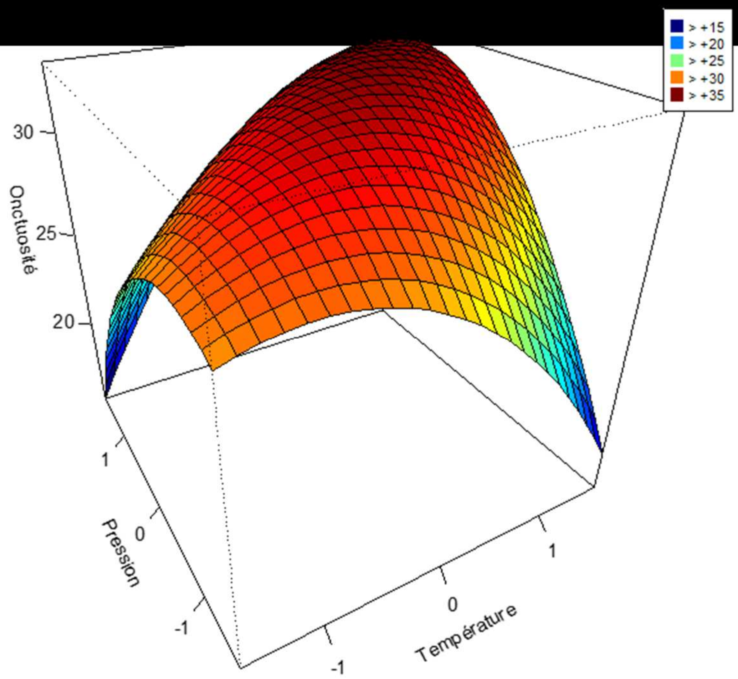
Recueil d'exercices sur les plans d'expériences



Les exercices présents dans ce recueil ont été construits par les étudiants de l'Institut Agro (Rennes) dans le cadre d'un module de planification expérimentale.

Ils sont le résultat de leur apprentissage sur ce cours, mais aussi de leur expérience acquise lors de stage. Certains exercices sont issus d'un contexte réel, avec des données réelles, d'autres exercices sont le fruit de leur imagination.

Exercices proposés par les élèves ingénieurs suivant le module de planification expérimentale

Enseignant responsable : François Husson

Institut Agro - Rennes

Juin 2020

Table des matières

Exercices sur les plans fractionnaires

Exercice 1. Construction d'un plan 2^{4-1} (par M. Loas, thématique réelle)

Nous souhaitons étudier la capacité de réhydratation d'une poudre suite à un séchage de produits alimentaires destiné à être introduit dans une soupe instantanée. Cette capacité de réhydratation dépend de 3 facteurs : la pression appliquée au cours du procédé de séchage, le temps de séchage et la teneur en eau initiale du produit avant séchage. Les facteurs étudiés présentent chacun 2 modalités :

Pression (bar)	2	5
Temps (min)	10	30
Teneur en eau (%)	12	20

1. **Combien d'essais sont nécessaires dans cette étude pour construire un plan complet ? Ecrire la matrice des essais et la matrice des effets du plan complet contenant tous les effets principaux et toutes les interactions.**

Il y a 3 facteurs à 2 modalités soit 2^3 essais, c'est-à-dire 8 essais.

La matrice des essais s'écrit :

Essais	Pression	Temps	Teneur en eau
1	5	30	20
2	5	30	12
3	5	10	20
4	5	10	12
5	2	30	20
6	2	30	12
7	2	10	20
8	2	10	12

En remplaçant, par convention, par des +1 et -1 :

Essais	Pression	Temps	Teneur en eau
1	+1	+1	+1
2	+1	+1	-1
3	+1	-1	+1
4	+1	-1	-1
5	-1	+1	+1
6	-1	+1	-1
7	-1	-1	+1
8	-1	-1	-1

La matrice des effets s'écrit :

I	Pression	Temps	Teneur en eau	Pression/Temps	Pression/ Teneur en eau	Temps/Teneur en eau	Pression/Temps/ Teneur en eau
1	+1	+1	+1	+1	+1	+1	+1
1	+1	+1	-1	+1	-1	-1	-1
1	+1	-1	+1	-1	+1	-1	-1
1	+1	-1	-1	-1	-1	+1	+1
1	-1	+1	+1	-1	-1	+1	-1
1	-1	+1	-1	-1	+1	-1	+1
1	-1	-1	+1	+1	-1	-1	+1
1	-1	-1	-1	+1	+1	+1	-1

2. **Pour des raisons d'approvisionnement, on doit réaliser l'étude avec des échantillons de produit provenant de deux tours de séchage différentes, une neuve et une plus ancienne. Comment interprétez-vous ce changement ? Que cela change-t-il par rapport aux conditions d'expérience ?**

Il faut ajouter un 4ème facteur à deux modalités à l'étude : le facteur « Tour de séchage » qui peut prendre la modalité « neuve » ou la modalité « ancienne ».

Le nombre d'essai augmente. Pour étudier un plan complet à 4 facteurs à 2 modalités, il faut $2^4 = 16$ essais.

3. **Nous sommes limités en nombre d'essais réalisables. On souhaite donc étudier tous les facteurs en 8 essais. Construire le plan sur R en utilisant la fonction FrF2 du package FrF2. Donner le ou les générateurs d'alias et la résolution du plan.**

La sortie R nous donne un générateur d'alias. Ce qui est cohérent car dans le cas d'un plan 2^4-1 il y a bien $2^1-1 = 1$ un générateur d'alias attendu.

Le plan construit est donc de résolution 4 car les effets principaux sont confondus avec des interactions d'ordre 3 ou plus.

```
library(FrF2)
plan <- FrF2(nruns=8, nfactors=4, factor.names =
list(Pression=c("basse","haute"),Temps=c("court","long"),TeneurH2O=c("min","max"),Tour=c("neuve"
,"ancienne")))
summary(plan)

Call:
FrF2(nruns = 8, nfactors = 4, factor.names = list(Pression = c("basse",
    "haute"), Temps = c("court", "long"), TeneurH2O = c("min",
    "max"), Tour = c("neuve", "ancienne")))

Experimental design of type  FrF2
8  runs

Factor settings (scale ends):
  Pression Temps TeneurH2O     Tour
1    basse court       min    neuve
2    haute  long       max ancienne

Design generating information:
$legend
[1] A=Pression  B=Temps     C=TeneurH2O D=Tour

$generators
[1] D=ABC

Alias structure:
$fi2
[1] AB=CD AC=BD AD=BC

The design itself:
  Pression Temps TeneurH2O     Tour
1    basse court       min    neuve
2    haute  long       max ancienne
3    haute court       min ancienne
4    basse  long       max    neuve
5    basse  long       min ancienne
6    haute  long       min    neuve
7    haute court       max    neuve
8    basse court       max ancienne
class=design, type= FrF2

X<-model.matrix(~.,data=plan)
t(X)%*%X

(Intercept) Pression1 Temps1 TeneurH2O1 Tour1
(Intercept)           8         0      0          0     0
Pression1             0         8      0          0     0
Temps1                0         0      8          0     0
TeneurH2O1            0         0      0          8     0
Tour1                 0         0      0          0     8
```

Exercice 2. Comment améliorer la texture d'un tartinable au saumon ? (exemple fictif, par Manon Filippi)

La conserverie de Poséidon souhaite commercialiser pour les fêtes de fin d'année son tout nouveau tartinable au saumon. Mais les essais industriels ne sont pas concluants, il y a trop d'exsudat dans la préparation. Le service R&D décide donc de retravailler la recette en étudiant 8 paramètres qui interviennent dans le procédé de fabrication :

- le poisson (cru, cuit)
- la quantité d'eau (7%, 10%)
- l'huile (colza, olive extra vierge)

- le jaune d'œuf liquide (bio, non bio)
- la quantité de jus de citron (2%,4%)
- la fécule (manioc, pomme de terre)
- crème fraîche (épaisse, semi-épaisse)
- temps d'agitation mécanique (6 min, 10min)

Les fêtes approchant à grand pas, l'entreprise ne peut réaliser qu'un nombre d'essais limités.

1. **Combien d'essais sont nécessaires pour construire le plan complet ?**

Pour étudier le plan complet à 8 facteurs à 2 niveaux, il faut $2^8 = 256$ essais.

2. **Quel plan de base choisissez-vous pour étudier les 8 facteurs en 32 essais ?**

Le plan de base est un plan complet en 32 essais, c'est donc le plan 2^5 (2^{8-3}).

3. **Donner les générateurs d'alias et la résolution du plan.**

Le plan 2^5 permet d'estimer les effets suivants :

I 1 2 3 4 5 12 13 14 15 23 24 25 34 35 45 123 124 125 134 135 145 234 235 245 345 1234 1235 1245 1345 2345 12345

On doit introduire 3 facteurs supplémentaires au plan de base pour avoir les 8 effets, on confond donc avec les interactions d'ordre le plus élevé -1. Par conséquent, on peut confondre par exemple le facteur 6 avec 1234, 7 avec 1235 et 8 avec 1245.
On obtient alors les générateurs d'alias initiaux suivants : I_a = 12346, I_b= 12357 et I_c=12458
On doit trouver 2^3-1 = 7 générateurs d'alias au total. Pour cela, on déduit les autres générateurs en faisant les produits des générateurs initiaux 2 à 2 puis le produit des 3.
On obtient ainsi les générateurs suivants : I_{ab}=4567 I_{bc}= 3478 I_{ac}= 3568 et I_{abc}=12678
La plus petite interaction avec laquelle la constante est confondue est d'ordre 4, le plan construit est donc de résolution IV.

4. **Peut-on estimer sans ambiguïté les effets principaux et les interactions d'ordre 2 ?**

Le plan étant de résolution IV, on peut estimer les effets principaux si on néglige les interactions d'ordre 3 et plus. Mais on ne pourra pas estimer les interactions d'ordre 2 car celles-ci sont confondues entre elles.

On a 9 paramètres à estimer (les 8 facteurs et la constante), comme on réalise un plan de résolution IV en 32 essais on pourra estimer 23 effets supplémentaires (32-9).

5. **Construire ce plan sous R et vérifier la qualité du plan.**

```
library(FrF2)
plan <- FrF2(nruns=32,nfactors=8,factor.names=list(poisson=c("cru","cuit"),
    eau=c("7","10"),huile=c("colza","olive"),oeuf=c("bio","conventionnel")
    ,citron=c("2","4"),fecule=c("manioc","pomme de terre"),
    creme=c("epaisse","semiepaisse"),agitation=c("6","10")))
summary(plan)
```

```
Call:
FrF2(nruns = 32, nfactors = 8, factor.names = list(poisson = c("cru",
    "cuit"), eau = c("7", "10"), huile = c("colza", "olive"),
    oeuf = c("bio", "conventionnel"), citron = c("2", "4"), fecule = c("manioc",
        "pomme de terre"), creme = c("epaisse", "semiepaisse"),
    agitation = c("6", "10")))

Experimental design of type FrF2
32  runs

Factor settings (scale ends):
  poisson eau huile          oeuf citron         fecule       creme agigation
1     cru   7 colza           bio      2         manioc     epaisse         6
2    cuit  10 olive conventionnel      4 pomme de terre semiepaisse        10

Design generating information:
$legend
[1] A=poisson   B=eau       C=huile     D=oeuf      E=citron    F=fecule
[7] G=creme     H=agigation

$generators
[1] F=ABC   G=ABD   H=ACDE

Alias structure:
$fi2
[1] AB=CF=DG  AC=BF     AD=BG     AF=BC     AG=BD     CD=FG     CG=DF
```

```
The design itself:
   Poisson  eau    huile          oeuf   citron          fecule      creme   agitation
1      cru  10 olive            bio      2         manioc semiepaisse        6
2     cuit  10 colza conventionnel      2         manioc semiepaisse       10
3      cru   7 colza            bio      2         manioc     epaisse       10
4     cuit   7 olive            bio      4         manioc semiepaisse        6
5      cru   7 colza conventionnel      4         manioc semiepaisse       10
6     cuit  10 colza conventionnel      4         manioc semiepaisse        6
7      cru   7 olive conventionnel      4 pomme de terre semiepaisse        6
8      cru   7 colza            bio      4         manioc     epaisse        6
9      cru   7 olive            bio      4 pomme de terre     epaisse       10
10    cuit  10 olive            bio      2 pomme de terre     epaisse       10
11    cuit   7 colza conventionnel      2 pomme de terre     epaisse       10
12     cru  10 colza conventionnel      2 pomme de terre     epaisse        6
13    cuit   7 olive conventionnel      4         manioc     epaisse       10
14     cru  10 olive conventionnel      4         manioc     epaisse        6
15     cru  10 colza            bio      4 pomme de terre semiepaisse        6
16     cru  10 olive            bio      4         manioc semiepaisse       10
17     cru  10 olive conventionnel      2         manioc     epaisse       10
18    cuit   7 olive conventionnel      2         manioc     epaisse        6
19     cru   7 olive            bio      2 pomme de terre     epaisse        6
20    cuit  10 colza            bio      2         manioc     epaisse        6
21     cru  10 colza            bio      2 pomme de terre semiepaisse       10
22    cuit  10 olive conventionnel      4 pomme de terre semiepaisse       10
23    cuit  10 olive conventionnel      2 pomme de terre semiepaisse        6
24    cuit   7 colza conventionnel      4 pomme de terre     epaisse        6
25     cru   7 colza conventionnel      2         manioc semiepaisse        6
26     cru  10 colza conventionnel      4 pomme de terre     epaisse       10
27    cuit   7 colza            bio      4 pomme de terre semiepaisse       10
28    cuit   7 colza            bio      2 pomme de terre semiepaisse        6
29     cru   7 olive conventionnel      2 pomme de terre semiepaisse       10
30    cuit   7 olive            bio      2         manioc semiepaisse       10
31    cuit  10 olive            bio      4 pomme de terre     epaisse        6
32    cuit  10 colza            bio      4         manioc     epaisse       10
class=design, type= FrF2
```

```
X <- model.matrix(~.,data=plan) # Constuire la matrice X
t(X)%*%X # Constuire la transposée de X
            (Intercept) poisson1 eau1 huile1 oeuf1 citron1 fecule1 creme1 agigation1
(Intercept)          32        0    0      0     0       0       0      0          0
poisson1              0       32    0      0     0       0       0      0          0
eau1                  0        0   32      0     0       0       0      0          0
huile1                0        0    0     32     0       0       0      0          0
oeuf1                 0        0    0      0    32       0       0      0          0
citron1               0        0    0      0     0      32       0      0          0
fecule1               0        0    0      0     0       0      32      0          0
creme1                0        0    0      0     0       0       0     32          0
agigation1            0        0    0      0     0       0       0      0         32
```

```
round(solve(t(X)%*%X),2) # inverser le produit de la transposée de X (X') et de X
                                    # fonction round pour arrondir à 2 chiffres après la virgule
        (Intercept) poisson1 eau1 huile1 oeuf1 citron1 fecule1 creme1 agigation1
(Intercept)         0.03     0.00 0.00   0.00  0.00    0.00    0.00   0.00       0.00
poisson1            0.00     0.03 0.00   0.00  0.00    0.00    0.00   0.00       0.00
eau1                0.00     0.00 0.03   0.00  0.00    0.00    0.00   0.00       0.00
huile1              0.00     0.00 0.00   0.03  0.00    0.00    0.00   0.00       0.00
oeuf1               0.00     0.00 0.00   0.00  0.03    0.00    0.00   0.00       0.00
citron1             0.00     0.00 0.00   0.00  0.00    0.03    0.00   0.00       0.00
fecule1             0.00     0.00 0.00   0.00  0.00    0.00    0.03   0.00       0.00
creme1              0.00     0.00 0.00   0.00  0.00    0.00    0.00   0.03       0.00
agigation1          0.00     0.00 0.00   0.00  0.00    0.00    0.00   0.00       0.03
```

La matrice est bien diagonale, il n'y a donc pas de confusion entre les facteurs.
La qualité du plan est donc vérifiée.

Exercice 3. Construction d'un plan 2^{5-2} (exemple réel, par Hector Chassagnon)

Dans le cadre de l'UC Agroécologie, nous avons dû réaliser une expérience de terrain qui consistait à quantifier et classer les vers de terre dans le sol. Les prélèvements se sont faits sur 2 sites géographiques différents en raison des 2 lieux de confinement des 2 , la Normandie et le Beaujolais.

L'objectif était d'étudier l'effet du contexte pédoclimatique (type de sol : limoneux ou argileux) et l'usage des sols (pratique conventionnelles ou BIO) sur les communautés lombriciennes. Chaque prélèvement devait être répété deux fois par modalités. Pour des raisons de temps nous ne pouvions faire que 4 prélèvements chacun soit 8 prélèvements au total.

NB : La pratique cultural BIO implique une fertilisation par de la matière organique, un travail du sol limité voir absent, une réduction voire suppression des traitements phytosanitaires et des rotations culturales incluant des prairies

Nous voulons étudier le nombre de vers de terre présents dans les sols, permettant d'améliorer la qualité du sol par leurs activités biologiques.

Pour cela nous étudions l'influence de facteurs qualitatifs, présentant chacun 2 modalités, sur une variable réponse Y correspondant au nombre de vers de terres comptés.

7 critères sont particulièrement étudiés :

- Usage du sol : Prairie – Culture
- Zone géographique : Beaujolais – Normandie
- Pratique culturale sur la parcelle : Conventionnelle – BIO
- Type de sol : Limoneux – argileux
- Répétition du prélèvement : 1 - 2

Cependant nous n'avons la possibilité de faire que 8 essais.

1. **Afin d'avoir que 8 essais, nous construisons le plan de base 2^{5-2}. Quels impacts entraîne le choix d'un tel plan ? Donner les générateurs d'alias de ce dernier.**

Le plan complet serait 2^5 = 32 essais

Le plan $2^{5-2} = 2^3 = 8$ permet de réduire considérablement le nombre d'essais. Cependant, ce type de plan entraîne des confusions puisque l'on doit ajouter 2 facteurs pour avoir les 5 effets ; ces derniers doivent donc être confondus avec des éléments que l'on peut considérer comme négligeable.

On doit obtenir $2^{\text{nb facteurs ajoutés}} - 1 = 2^2 - 1 = 3$ générateurs

On confond les deux derniers facteurs (facteur 4, type de sol et facteur 5, répétition du prélèvement) avec des interactions d'ordre le plus élevé -1.

Le facteur 4 est alors confondu avec l'interaction 12 et le facteur 5 avec 23.

On a alors les générateurs d'alias suivant : I = 124 = 235 = 1345

2. **Quelle est la résolution du plan ?**

Le plan construit ici est de résolution 3. Ce chiffre correspond à l'ordre de l'interaction la plus petite confondue avec la constante I.

3. **Avec un tel plan, peut-on estimer sans ambiguïté tous les effets principaux et les interactions d'ordre 2 ?**

Avec un plan de résolution 3 on ne peut pas estimer les interactions mais seulement les effets principaux en faisant l'hypothèse que les interactions d'ordre 2 ou plus sont négligeables.

A travers cet exercice on s'intéresse à 5 facteurs. Cela nécessite l'estimation de 6 paramètres puisqu'il y a aussi 1 paramètre pour la constante I.

Nous effectuons 8 essais : il nous reste donc 2 degrés de liberté disponibles pour l'estimation de 2 interactions d'ordre 2.

4. **Que faudrait-il faire pour pouvoir estimer sans ambiguïté tous les facteurs principaux et toutes les interactions d'ordre 2 ? Quelles hypothèses doit-on faire dans ce cas ?**

Pour pouvoir estimer toutes les interactions on doit construire un plan avec plus d'essais. Pour cela il faudrait passer à un plan de résolution IV et faire 16 essais ; les effets principaux seraient donc confondus avec des interactions d'ordre 3 ou plus. Il faut alors faire l'hypothèse que ces interactions sont négligeables : cette hypothèse est raisonnable.

5. **Comment peut-on analyser les résultats ?**

Pour analyser les résultats de ce plan nous devons faire une analyse de variance à 5 facteurs (correspondant aux 5 critères). Il ne faudra pas oublier de prendre en compte la résiduelle.

Exercice 4. Recette de pancakes (exemple réel, par Léa Pautrel)

Le but de cette expérience est de trouver la recette de pancakes la plus appréciée. Voici les six variables à deux modalités qui vont être testées :

Facteur	Code	Modalité 1 $(= 1)$	Modalité 2 $(= -1)$
Type de farine	A	Farine blanche T65	Farine complète T150
Type de sucre	B	Sucre blanc	Sucre muscovado complet

Présence de banane	C	Avec banane	Sans banane
Présence d'épices	D	Cannelle + Muscade	Sans épices
Liant	E	Œufs	Graines de chia
Matière grasse (cuisson)	F	Beurre	Huile de tournesol

Deux pancakes seront réalisés par recette, afin que deux juges testent chacune des recettes. Afin d'éviter de manger des quantités astronomiques de pancakes, seules 8 recettes seront testées.

1. **Combien d'essais sont nécessaires pour construire un plan complet ?**

Il s'agit d'un plan de 6 facteurs à 2 modalités. Le plan complet nécessite donc $2^6 = 64$ essais.

2. **Donner le plan de base et la matrice des effets de ce plan.**

Il s'agit du plan fractionnaire 2^{6-3}. Le plan de base en 8 essais est le plan 2^3. Ce plan permet d'estimer les effets suivants :

I	A	B	C	AB	AC	BC	ABC

La matrice des effets de ce plan 2^3 est la suivante :

I	A	B	C	AB	AC	BC	ABC
1	1	1	1	1	1	1	1
1	1	1	-1	1	-1	-1	-1
1	1	-1	1	-1	1	-1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	1	-1	-1	1	-1
1	-1	1	-1	-1	1	-1	1
1	-1	-1	1	1	-1	-1	1
1	-1	-1	-1	1	1	1	-1

3. **Déterminer les confusions d'effets principaux, d'interactions d'ordre 2 et d'ordre 3 et les générateurs d'alias. En déduire la résolution du plan.**

Il faut ajouter 3 facteurs au plan de base pour estimer les 6 effets principaux. Ces facteurs seront confondus avec les interactions d'ordre 2 ($ordre\ le\ plus\ élevé\ - 1$) car plus d'un facteur doit être introduit. On confond alors D avec AB, E avec AC et F avec BC.

Les confusions sont les suivantes (seuls les effets principaux et les interactions de 2nd et 3ème ordre sont représentés) :

I	A	B	C	D	E	F	ABC
ABD	BD	AD	AE	AB	AC	BC	AF
ACE	CE	CF	BF	EF	DF	DE	BE
BCF	BEF	AEF	ADF	ACF	ABF	ABE	CD
DEF	CDF	CDE	BDE	BCE	BCD		ADE
							BDF
							CEF

Les calculs réalisés afin de trouver ces confusions sont ci-dessous :

$$AD = AAB = B$$
$$AE = AAC = C$$
$$AF = ABC$$
$$BD = BAB = A$$
$$BE = ABC$$
$$BF = BBC = C$$
$$CD = ABC$$
$$CE = CAC = A$$

$$CF = CBC = B$$
$$DE = ABAC = BC$$
$$DF = ABBC = AC$$
$$EF = ACBC = AB$$
$$ABD = I$$
$$ABE = ABAC = BC$$
$$ABF = ABBC = AC$$
$$ACD = ACAB = CB$$

$$ACE = I$$
$$ACF = ACBC = AB$$
$$ADE = AABAC = ABC$$
$$ADF = AABBC = C$$
$$AEF = AACBC = B$$
$$BCD = BCAB = AC$$
$$BCE = BCAC = AB$$
$$BCF = I$$

$$BDE = BABAC = C$$
$$BDF = BABBC = ABC$$
$$BEF = BACBC = A$$
$$CDE = CABAC = B$$
$$CDF = CABBC = A$$
$$CEF = CACBC = ABC$$
$$DEF = ABACBC = I$$

Il y a $2^3 - 1 = 7$ générateurs d'alias, qui sont donc les suivants :

$$I = ABD = ACE = BCF = DEF = ABEF = ACDF = BCDE$$

On pourra estimer uniquement les effets principaux. On émet l'hypothèse que les interactions d'ordre 2 ou plus sont négligeables. Il s'agit d'un plan de résolution III.

4. **Peut-on estimer quelques interactions d'ordre 2 avec ce plan d'expérience ? Si oui, combien ?**

8 essais sont proposés et vont permettre de déterminer 7 paramètres : la constante I et les 6 effets principaux (A, B, C, D, E, F). Il reste un degré de liberté disponible. On peut donc estimer sans ambiguïté un autre paquet d'effet. Grâce au tableau des confusions ci-dessus, on trouve que l'on peut estimer l'une des interactions AF, BE ou CD si l'on suppose que les deux autres sont négligeables. Si l'on reprend le tableau décrivant les facteurs à tester, on pourrait donc potentiellement estimer les interactions suivantes :

- $Type\ de\ farine\ (A) \times Matière\ grasse\ (F)$
- $Type\ de\ sucre\ (B) \times Liant\ (E)$
- $Présence\ de\ banane\ (C) \times Présence\ d'épices\ (D)$

Cependant, il n'y a pas a priori de raison de penser que l'une de ces interactions est plus importante que les deux autres. Aucune interaction ne sera donc estimée par la suite.

5. **Construire ce plan sous R en utilisant la fonction FrF2 du package FrF2. Calculer la matrice $(X'X)^{-1}$ avec les effets linéaires uniquement (utiliser la fonction model.matrix pour calculer X). Refaire la même chose en ajoutant au modèle les interactions de deuxième ordre. Interpréter.**

```
library(FrF2)
plan <- FrF2(nruns=8, nfactors=6)
summary(plan)

Call:
FrF2(nruns = 8, nfactors = 6)

[. . .]

Design generating information:
$legend
[1] A=A B=B C=C D=D E=E F=F

$generators
[1] D=AB E=AC F=BC

Alias structure:
$main
[1] A=BD=CE B=AD=CF C=AE=BF D=AB=EF E=AC=DF F=BC=DE

$fi2
[1] AF=BE=CD

The design itself:
   A  B  C  D  E  F
1  1  1  1  1  1  1
2 -1 -1  1  1 -1 -1
3  1 -1 -1 -1 -1  1
4 -1  1  1 -1 -1  1
5 -1 -1 -1  1  1  1
6  1 -1  1 -1  1 -1
7  1  1 -1  1 -1 -1
8 -1  1 -1 -1  1 -1
class=design, type= FrF2

# Analyse de la qualité de ce plan
options(contrasts=c("contr.sum","contr.sum"))
X <- model.matrix(~.,data=plan)
t(X)%*%X
            (Intercept) A1 B1 C1 D1 E1 F1
(Intercept)           8  0  0  0  0  0  0
A1                    0  8  0  0  0  0  0
B1                    0  0  8  0  0  0  0
C1                    0  0  0  8  0  0  0
D1                    0  0  0  0  8  0  0
E1                    0  0  0  0  0  8  0
F1                    0  0  0  0  0  0  8
```

Le code ainsi que les résultats obtenus pour les effets linéaires sont présentés ci-dessus. Les résultats sont bien identiques à ceux obtenus par avant, et les effets principaux sont bien orthogonaux.

```
X <- model.matrix(~ A + B + C + D + E + F + A:B + A:C + A:D + A:E + A:F + B:C + B:D + B:E + B:F
+ C:D + C:E + C:F + D:E + D:F + E:F, data=plan)
t(X)%*%X
```

```
            (Intercept) A1 B1 C1 D1 E1 F1 A1:B1 A1:C1 A1:D1 A1:E1 A1:F1 B1:C1
(Intercept)           8  0  0  0  0  0  0     0     0     0     0     0     0
A1                    0  8  0  0  0  0  0     0     0     0     0     0     0
B1                    0  0  8  0  0  0  0     0     0     8     0     0     0
C1                    0  0  0  8  0  0  0     0     0     0     8     0     0
D1                    0  0  0  0  8  0  0     8     0     0     0     0     0
E1                    0  0  0  0  0  8  0     0     8     0     0     0     0
F1                    0  0  0  0  0  0  8     0     0     0     0     0     8
A1:B1                 0  0  0  0  8  0  0     8     0     0     0     0     0
A1:C1                 0  0  0  0  0  8  0     0     8     0     0     0     0
A1:D1                 0  0  8  0  0  0  0     0     0     8     0     0     0
A1:E1                 0  0  0  8  0  0  0     0     0     0     8     0     0
A1:F1                 0  0  0  0  0  0  0     0     0     0     0     8     0
B1:C1                 0  0  0  0  0  0  8     0     0     0     0     0     8
B1:D1                 0  8  0  0  0  0  0     0     0     0     0     0     0
B1:E1                 0  0  0  0  0  0  0     0     0     0     0     8     0
B1:F1                 0  0  0  8  0  0  0     0     0     0     8     0     0
C1:D1                 0  0  0  0  0  0  0     0     0     0     0     8     0
C1:E1                 0  8  0  0  0  0  0     0     0     0     0     0     0
C1:F1                 0  0  8  0  0  0  0     0     0     8     0     0     0
D1:E1                 0  0  0  0  0  0  8     0     0     0     0     0     8
D1:F1                 0  0  0  0  0  8  0     0     8     0     0     0     0
E1:F1                 0  0  0  0  8  0  0     8     0     0     0     0     0
            B1:D1 B1:E1 B1:F1 C1:D1 C1:E1 C1:F1 D1:E1 D1:F1 E1:F1
(Intercept)     0     0     0     0     0     0     0     0     0
A1              8     0     0     0     8     0     0     0     0
B1              0     0     0     0     0     8     0     0     0
C1              0     0     8     0     0     0     0     0     0
D1              0     0     0     0     0     0     0     0     8
E1              0     0     0     0     0     0     0     8     0
F1              0     0     0     0     0     0     8     0     0
A1:B1           0     0     0     0     0     0     0     0     8
A1:C1           0     0     0     0     0     0     0     8     0
A1:D1           0     0     0     0     0     8     0     0     0
A1:E1           0     0     8     0     0     0     0     0     0
A1:F1           0     8     0     8     0     0     0     0     0
B1:C1           0     0     0     0     0     0     8     0     0
B1:D1           8     0     0     0     8     0     0     0     0
B1:E1           0     8     0     8     0     0     0     0     0
B1:F1           0     0     8     0     0     0     0     0     0
C1:D1           0     8     0     8     0     0     0     0     0
C1:E1           8     0     0     0     8     0     0     0     0
C1:F1           0     0     0     0     0     8     0     0     0
D1:E1           0     0     0     0     0     0     8     0     0
D1:F1           0     0     0     0     0     0     0     8     0
E1:F1           0     0     0     0     0     0     0     0     8
```

Une analyse avec les interactions de 2nd ordre (code ci-dessus) confirme que les interactions d'ordre 2 ne peuvent pas être analysés (non orthogonaux).

6. Comment analyser les résultats obtenus ?

J'ai réalisé les expériences chez moi afin de pouvoir analyser les résultats. Le plan d'expérience ainsi que les résultats sont présentés dans le tableau suivant. La note correspond à la moyenne de la note des deux juges.

	Farine	*Sucre*	*Banane*	*Epices*	*Liant*	*Gras*	*Note*
1	T65	Blanc	Sans	Avec	Graines de chia	Huile de tournesol	3
2	T150	Blanc	Sans	Sans	Œuf	Huile de tournesol	5

3	T65	Muscovado	Avec	Sans	Œuf	Huile de tournesol	9
4	T150	Muscovado	Avec	Avec	Graines de chia	Huile de tournesol	10
5	T65	Blanc	Avec	Avec	Œuf	Beurre	8
6	T150	Muscovado	Sans	Avec	Œuf	Beurre	6
7	T150	Blanc	Avec	Sans	Graines de chia	Beurre	7
8	T65	Muscovado	Sans	Sans	Graines de chia	Beurre	4

Il faut réaliser une analyse de variance avec la variable réponse Y la note et 6 variables explicatives correspondant aux différents facteurs. Il n'y a pas assez de degrés de liberté disponibles pour estimer les interactions. Le modèle est donc le suivant :

$$Y_{ijklmn} = \mu + A_i + B_j + C_k + D_l + E_m + F_n + \varepsilon_{ijklmn}$$

Avec A_i l'effet de la farine, B_j l'effet du sucre, C_k l'effet de la banane, D_l l'effet des épices, E_m l'effet du liant, F_n l'effet du gras et ε_{ijklmn} la résiduelle.

```
# Importation des données
pancake = read.csv("_Exam-data.csv", header = TRUE, sep = ";", encoding = "utf-8")
summary(pancake) # Vérification que l'importation a bien fonctionné
```

```
Farine        Sucre     Banane    Epices              Liant                    Gras            Note
T150:4   Blanc    :4   Avec:4   Avec:4   Graines de chia:4   Beurre            :4   Min.   : 3.00
T65 :4   Muscovado:4   Sans:4   Sans:4   Œuf            :4   Huile de tournesol:4   1st Qu.: 4.75
                                                                                    Median : 6.50
                                                                                    Mean   : 6.50
                                                                                    3rd Qu.: 8.25
                                                                                    Max.   :10.00
```

```
mod <-lm(Note~Farine+Sucre+Banane+Epices+Liant+Gras, data=pancake) # Modèle linéaire
summary(mod)
```

```
Call:
lm(formula = Note ~ Farine + Sucre + Banane + Epices + Liant + Gras, data = pancake)

Residuals:
    1     2     3     4     5     6     7     8 
-0.25  0.25 -0.25  0.25  0.25 -0.25 -0.25  0.25 

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)  
(Intercept)              7.7500     0.6614   11.72   0.0542 .
FarineT65               -1.0000     0.5000   -2.00   0.2952  
SucreMuscovado           1.5000     0.5000    3.00   0.2048  
BananeSans              -4.0000     0.5000   -8.00   0.0792 .
EpicesSans              -0.5000     0.5000   -1.00   0.5000  
LiantŒuf                 1.0000     0.5000    2.00   0.2952  
GrasHuile de tournesol   0.5000     0.5000    1.00   0.5000  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7071 on 1 degrees of freedom
Multiple R-squared:  0.9881,	Adjusted R-squared:  0.9167 
F-statistic: 13.83 on 6 and 1 DF,  p-value: 0.203
```

Il n'y a donc pas d'effet significatif d'aucun des facteurs au seuil de 5 %. Cependant, la banane semble être le facteur ayant l'effet le plus important. Avec une p-value de 0.08, pas si éloignée de 0.05, on pourrait estimer que l'absence de banane a donc un effet négatif sur l'appréciation (ou, qu'au contraire, la présence de banane a un effet positif.)

Peut-être aurait-on pu considérer le juge comme un 7ème effet ? Le plan aurait alors pu comporter 16 essais, chaque essai correspondant à une dégustation, et non plus à une recette.

Exercice 5. Recette de cookies (exemple fictif, par Zoé Wante)

On étudie l'influence de cinq facteurs à deux modalités sur l'appréciation d'un cookie. Les facteurs sont les suivants : la température du four (180ºC ou 200 ºC), le type des pépites (formées industriellement ou chocolat en morceau), le temps de cuisson (8 min ou 12min), le type de chocolat (noir à 70% ou 80% de cacao), vanille (avec ou sans). Nous considèrerons qu'une seule interaction n'est pas négligeable : température - temps de cuisson.

Ces cookies seront dégustés par des juges. Afin que les juges puissent noter correctement les cookies ils n'en dégusteront pas plus de 8. Et pour que l'effet séance n'intervienne pas une seule séance de dégustation aura lieu.

1. **Proposez un plan en 8 essais, détaillez votre démarche.**

Pour réaliser un plan complet il aurait fallu $2^5 = 32$ essais or ici nous ne disposons que de 8 essais possibles. Pour éviter les confusions d'effet il faut donc construire un plan d'expérience.

Nous avons construit le tableau suivant :

	-1	1
Température	180	200
Vanille	Avec	Sans
Pépite	Industrielles	Morceau
Type	70%	80%
Temps	8 min	12 min

De manière informatique :

```
Variables <- matrix(c("180","200","avec","sans","industrielle","morceau","70%","80%","8
min","12min"), nrow = 5, ncol = 2, byrow= TRUE, dimnames = list(c("Températue","Vanille
","Pépite","Type","Temps"),c("-1","1")))
```

Nous allons réaliser un plan fractionnaire 2^{5-2}. Il faudra faire attention au moment du choix des confusions d'ordre 2 de ne pas confondre l'interaction température:temps avec des effets d'ordre 1. Nous construisons donc un plan complet 2^3 à la main que nous affichons sur R :

1	-1	-1	1
2	-1	1	1
3	1	-1	1
4	-1	-1	-1
5	1	1	1
6	1	1	-1
7	1	-1	-1
8	_-1	1	-1

```
plan3 <- matrix(c(-1,-1,1,-1,1,1,1,-1,1,-1,-1,-1,1,1,1,1,1,-1,1,-1,-1,-1,1,-1), nrow =
8, ncol = 3, byrow= TRUE, dimnames = list(c("1","2","3","4","5","6","7","8"),c("F1","F2
","F3")))
```

Nous choisissons désormais les générateurs d'alias. On pose pour F4 : 4 = 13 soit I = 134, on confond ici l'effet de F4 avec l'interaction F1F3 qui est, selon l'énoncé, négligeable. Et on pose F5 : 5 = 23 soit I = 235. On confond l'effet de F5 avec l'interaction F2F3 qui est aussi négligeable. On regarde alors l'ensemble des générateurs d'alias :

	1	2	3	12	13	23
I = 134	34	1234	14	234	4	124
I = 235	1235	35	25	135	125	5

I = 1245	245	145	12345	45	2345	1345

```
alias <- matrix(c(34, 1234,14,234,4,124,1235,35,25,135,125,5,245,145,12345,45,2345,1345
), nrow = 3, ncol = 6, byrow= TRUE, dimnames = list(c("I = 134","I = 235", "I = 1245"),
c("1","2","3","12","13","23"))
```

On observe que les effets principaux sont confondus avec des interactions d'ordre 2 on a donc une résolution III et l'interaction température:temps est confondue avec des interactions d'ordre 2 ou plus qui sont supposées négligeables. Cette interaction pourra donc être étudiée.

Nous construisons donc le plan fractionnaire suivant :

```
   A  B  C  D  E
1 -1 -1  1  1 -1
2 -1  1  1 -1 -1
3  1 -1  1 -1  1
4 -1 -1 -1  1  1
5  1  1  1  1  1
6  1  1 -1  1 -1
7  1 -1 -1 -1 -1
8 -1  1 -1 -1  1
```

```
set.seed(2)
library(FrF2)
planfinal <- FrF2(nfactors=5, resolution=3)
```

Nous obtenons donc un plan fractionnaire 2^{5-2} de résolution III dont les effets principaux sont confondus avec les effets d'ordre 2 ou plus. Les effets principaux sont orthogonaux deux à deux puisque chaque modalité est présente autant de fois pour chaque facteur. L'interactions d'ordre 2 : F1F2 peut être étudiée. Enfin le modèle n'est pas saturé : ddl = 8.

Exercice 6. Quelle méthode de préparation d'échantillons est la meilleure ? (exemple réel, par Léa Carrolaggi)

L'énoncé de cet exercice repose sur une différence de méthode réelle observée dans deux entreprises d'un même groupe agroalimentaire, l'une se situant en France et l'autre en Angleterre. Cette méthode concerne la préparation d'échantillons de camembert à analyser d'un point de vue physico-chimique à l'aide d'un Food Scan. Le Food Scan est un appareil de spectrométrie infrarouge qui est utilisé dans ces entreprises pour entre autres déterminer de manière très rapide l'extrait sec des camemberts à différentes étapes du processus de fabrication pour maîtriser leur cinétique d'égouttage et d'acidification. Le fournisseur de cet appareil ne procure pas d'informations précises sur la préparation de l'échantillon à analyser, il indique seulement que l'échantillon doit être disposé dans une boîte de Pétrie. Il semblerait que la précision de mesure d'extrait sec dépende principalement de 5 facteurs :

- le mode de préparation de l'échantillon (noté A) : mixeur ou moulin râpe
- le mode de remplissage de la boîte de Pétrie (noté B) : spatule ou cuillère
- la quantité de matière (noté C) : boîte remplie de moitié ou entièrement remplie
- la surface de l'échantillon (noté D) : lisse ou biscornue
- l'aspect de l'échantillon (noté E) : présence de trous d'air ou absence de trous d'air

On souhaite donc étudier l'influence de ces 5 facteurs sur la précision de mesure d'extrait sec. Seulement, la problématique est urgente et ne doit pas nécessiter un grand coût de mise en œuvre, c'est pourquoi il sera raisonnable de se limiter à 8 expériences.

1. **Quel est le nom du plan à construire ?**

On s'intéresse ici à l'effet de 5 variables explicatives qualitatives à deux modalités sur une variable à expliquer quantitative. Le plan à construire est donc le plan fractionnaire 2^{5-2} en 8 essais.

2. **Construire ce plan en suivant la méthodologie vue en cours (choix d'un plan de base, construction de la matrice des effets du modèle saturé associé à ce plan de base, choix des confusions, détermination des confusions résultantes).**

Le plan de base que l'on va choisir est un plan complet en 8 essais, c'est donc le plan 2^3 suivant:

A	B	C
1	1	1
1	1	-1
1	-1	1
1	-1	-1
-1	1	1
-1	1	-1
-1	-1	1
-1	-1	-1

La matrice des effets contenant tous les effets principaux, toutes les interactions d'ordre 2 et l'interaction d'ordre 3 associés au plan de base 2^3 est la suivante :

I	A	B	C	AB	AC	BC	ABC
1	1	1	1	1	1	1	1
1	1	1	-1	1	-1	-1	-1
1	1	-1	1	-1	1	-1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	1	-1	-1	1	-1
1	-1	1	-1	-1	1	-1	1
1	-1	-1	1	1	-1	-1	1
1	-1	-1	-1	1	1	1	-1

Il va maintenant falloir confondre les facteurs D et E que l'on veut ajouter dans le modèle avec des interactions que l'on considère négligeables. Il n'est bien évidemment pas possible de confondre ces facteurs avec des facteurs principaux car on n'arriverait pas à savoir ce qui est dû uniquement des facteurs principaux de ce qui est dû des autres. Ici on souhaite ajouter deux facteurs donc on va les confondre avec des interactions d'ordre les plus élevées – 1 à savoir : AB et BC (on aurait très bien pu les confondre avec AB et AC ou AC et BC). Voici toutes les confusions résultantes :

I	A	B	C	AB	AC	BC	ABC
ABD	BD	AD	ABCD	D	BCD	ACD	CD
BCE	ABCE	CE	BE	ACE	ABE	E	AE
ACDE	CDE	ABCDE	ADE	BCDE	DE	ABDE	BDE

3. **Quelle est la qualité du plan construit ?**

Le plan construit est de résolution III car la plus petite interaction avec laquelle la constante est confondue est d'ordre 3 (l'ordre le moins élevé parmi tous les générateurs d'alias est 3) et les effets principaux sont confondus avec des interactions d'ordre 2 ou plus. En résolution III, il est possible d'estimer sans ambiguïté les effets principaux si on considère que toutes les interactions d'ordre 2 ou plus sont négligeables. Cependant, il ne sera pas possible d'estimer sans ambiguïté les interactions d'ordre 2 car certaines d'entre elles sont confondues avec d'autres interactions d'ordre 2.

4. **Quel est le modèle qui permettra d'analyser les résultats à l'issue des expériences ?**

C'est le modèle d'analyse de variance à 5 facteurs. Le plan étant de résolution III, on ne peut pas étudier le modèle avec les effets d'interactions mais seulement le modèle avec les effets principaux. Il s'écrit de la manière suivante :

$$\forall(i,j,k,l,m)\quad Y_{ijklm} = \mu+\alpha_i+\beta_j+\gamma_k+\delta_l+\kappa_m+\varepsilon_{ijklm}$$

Avec α_i l'effet du mode de préparation de l'échantillon, β_j l'effet du mode de remplissage de la boîte de Pétrie, γ_k l'effet de l'aspect de l'échantillon, δ_l l'effet de la surface de l'échantillon, κ_m l'effet de la quantité de matière et ε_{ijklm} les résidus sur lesquels on émet les hypothèses suivantes :

$$\forall\ (i, j, k, l, m)\ L(\varepsilon_{ijklm}) = N\ (0, \sigma)$$
$$\forall\ (i, j, k, l, m) \neq (i', j', k', l', m')$$
$$\mathrm{cov}(\varepsilon_{ijklm}, \varepsilon_{i'j'k'l'm'}) = 0$$

5. **Construire le plan sous R en précisant le nom des facteurs ainsi que leurs modalités dans le même ordre que celui donné dans l'énoncé et en fixant le paramètre seed de la fonction que vous utiliserez à 123456. Une fois le plan construit, indiquer la première expérience que l'expérimentateur devra réaliser.**

```
library(FrF2) #Chargement du package FrF2 pour construire des plans fractionnaires
planfractionnaire <- FrF2(nruns=8, nfactors=5,
factor.names=list(mode_prepa=c("mixeur","moulin rape"), mode_remp=c("spatule",'cuillere'),
qtite_mat=c("moitie remplie","entierement remplie"), surface_ech=c("lisse","biscornue"),
aspect_ech=c("trous d air","pas trous d air")),seed=123456) #Construction du plan fractionnaire
25-2, le nombre d'essais est précisé dans nruns, le nombre de facteurs est indiqué dans nfactors
et les noms des facteurs ainsi que leurs modalités sont répertoriés dans factor.names, seed
permet de fixer la configuration du plan d'expérience
planfractionnaire #Affichage du plan fractionnaire construit
   mode_prepa mode_remp           qtite_mat surface_ech      aspect_ech
1 moulin rape  cuillere      moitie remplie   biscornue     trous d air
2 moulin rape   spatule      moitie remplie       lisse     trous d air
3      mixeur  cuillere entierement remplie       lisse     trous d air
4      mixeur   spatule      moitie remplie   biscornue pas trous d air
5 moulin rape   spatule entierement remplie       lisse pas trous d air
6      mixeur  cuillere      moitie remplie       lisse pas trous d air
7      mixeur   spatule entierement remplie   biscornue     trous d air
8 moulin rape  cuillere entierement remplie   biscornue pas trous d air
class=design, type= FrF2
```

D'après le plan renvoyé par R, pour la première expérience, un échantillon devra être préparé à l'aide d'un moulin râpe puis déposé dans une boîte de Pétrie à l'aide d'une cuillère en veillant à ce que la boîte soit à moitié remplie, la surface soit biscornue et le contenu présente des trous d'air.

6. **Pourquoi n'est-il pas nécessaire de construire la matrice $(X'X)^{-1}$ associé à ce plan pour tester l'orthogonalité entre les facteurs principaux ?**

Il n'est pas nécessaire de construire la matrice $(X'X)^{-1}$ puisque comme dit précédemment, dans un plan fractionnaire de résolution III, les facteurs principaux sont estimés sans ambiguïté dans la mesure où on considère que les interactions d'ordre 2 sont négligeables. Il y a donc orthogonalité entre les facteurs principaux.

7. **Il semblerait que l'interaction AE ne soit pas négligeable (l'aspect de l'échantillon semble étroitement dépendre de son mode de préparation). Il paraît donc tentant de rajouter cette interaction dans notre modèle d'analyse de variance. Construire la matrice X associée à ce plan et au modèle comprenant les effets linéaires et l'effet de l'interaction AE puis calculer la matrice $(X'X)^{-1}$. Interpréter.**

```
X <-model.matrix(~mode_prepa+mode_remp+qtite_mat+surface_ech+aspect_ech
+mode_prepa:aspect_ech,data=planfractionnaire) #Création de la matrice X des effets principaux
et de l'effet d'interaction
t(X)%*%X #Affichage de la matrice X'X
           (Intercept) mode_prepa1 mode_remp1 qtite_mat1 surface_ech1 aspect_ech1 mode_prepa1:aspect_ech1
(Intercept)          8           0          0          0            0           0                       0
mode_prepa1          0           8          0          0            0           0                       0
mode_remp1           0           0          8          0            0           0                       0
qtite_mat1           0           0          0          8            0           0                       8
surface_ech1         0           0          0          0            8           0                       0
aspect_ech1          0           0          0          0            0           8                       0
mode_prepa1:asct_ech1 0          0          0          8            0           0                       8
```

On se retrouve avec une matrice X'X qui pose problème car deux colonnes sont exactement les mêmes à cause d'une confusion totale entre le facteur C et le facteur d'interaction AE, ce qui la rend non inversible. C'est ce que témoigne le message d'erreur renvoyé lors de l'exécution de la ligne de code permettant d'inverser cette matrice X'X :

```
solve(t(X)%*%X) #Affichage de la matrice (X'X)-1
Error in solve.default(t(X) %*% X) :
  routine Lapack dgesv : le système est exactement singulier : U[7,7] = 0
```

En intégrant l'effet de l'interaction AE dans notre modèle d'analyse de variance, il nous est impossible de dissocier ce qui est dû du facteur C de ce qui est dû du facteur AE. Si l'interaction AE n'est vraiment pas négligeable, il n'y aura pas d'autres solutions que de faire plus d'essais. En 16 essais, il sera possible de construire un plan fractionnaire de résolution 4 dans lequel les effets principaux seront estimés sans ambigüité en considérant que les interactions d'ordre 3 ou plus sont négligeables et l'effet d'interaction AE pourra être estimée sans ambiguïté en considérant que les autres interactions d'ordre 2 sont négligeables.

8. **Le facteur quantité de matière aurait pu être à 4 modalités : boîte remplie au ¼ ou boîte remplie au ½ ou boîte remplie au ¾ ou boîte entièrement remplie. Dans ce cas-là quel plan auriez-vous construit et comment l'auriez-vous construit ?**

On se serait intéressé ici à l'effet de 5 variables explicatives qualitatives (4 à deux modalités et 1 à quatre modalités) sur une variable à expliquer quantitative. On reconnait donc un plan asymétrique $L_8\ 2^{4-1}\ 4$. Pour construire un tel plan on aurait commencé par construire le plan fractionnaire 2^{6-3} en 8 essais puis on aurait utilisé deux facteurs à 2 niveaux pour construire le facteur E à 4 niveaux (les 4 combinaisons possibles seraient recodées comme 4 modalités).

Exercice 7. Lavage de t-shirts personnalisés faits maison (thématique réelle, par M. Madrolle)

La thématique de cet exercice est réelle, mais les expériences n'ont pas été effectuées.

Afin de renouveler sa garde-robe, une personne cherche à personnaliser d'anciens t-shirts blancs en peignant dessus différents dessins. Elle souhaite étudier la tenue de la peinture après plusieurs lavages, le but étant de les porter le plus possible. Selon elle, la tenue de la peinture (notée sur une échelle de 0 à 10) peut dépendre de différents facteurs à 2 modalités :

- Qualité de la couche : « Fine » ou « Épaisse »
- Mode de lavage : « Doux » ou « Normal »
- Type de peinture : « Acrylique » ou « Posca »
- Tissu du t-shirt : « Coton » ou « Polyester »
- Protection : lavage dans une « housse » ou « sans housse »

L'objectif est d'établir un plan d'expérience pour étudier la tenue de la peinture dans le temps selon la méthode de lavage et de peignage.

1. **Combien d'essais sont nécessaires pour construire un plan complet à 5 facteurs ?**

Pour étudier un plan complet à 5 facteurs à 2 niveaux, il faut $2^5 = 32$ essais.

2. **Quel plan de base choisissez-vous pour effectuer tous les facteurs en 8 essais ?**

Le plan de base est un plan complet en 8 essais, c'est donc le plan $2^{5-2} = 2^3 = 8$ essais.

3. **Construire la matrice des effets du modèle saturé avec ce plan de base. Déterminez toutes les interactions, et en déduire la résolution du plan.**

La matrice des effets s'écrit donc :

				4	5		
I	1	2	3	12	13	23	123
1	1	1	1	1	1	1	1
1	1	1	-1	1	-1	-1	-1
1	1	-1	1	-1	1	-1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	1	-1	-1	1	-1
1	-1	1	-1	-1	1	-1	1
1	-1	-1	1	1	-1	-1	1
1	-1	-1	-1	1	1	1	-1

On ajoute 2 facteurs, ce qui fait qu'il va y avoir des confusions avec les interactions d'ordre le plus élevé -1. Ainsi, le facteur 4 est confondu avec l'interaction 12 et 5 avec l'interaction 13.

4. **Saisir les données puis faire l'analyse avec R. Donner les générateurs d'alias et la résolution de ce plan.**

```
#Construction du plan
library(FrF2)
```

```
plan<-
FrF2(nruns=8,nfactors=5,factor.names=list(couche=c("epaisse","fine"),lavage=c("doux","normal"),
peinture=c("acrylique","posca"),tissu=c("coton","polyester"),protection=c("oui","non")))
summary(plan)

Call :
FrF2(nruns = 8, nfactors = 5, factor.names = list(couche = c("epaisse",
    "fine"), lavage = c("doux", "normal"), peinture = c("acrylique",
    "posca"), tissu = c("coton", "polyester"), protection = c("oui",
    "non")))

Experimental design of type  FrF2
8  runs #nombre d'essais à réaliser

Factor settings (scale ends):

Design generating information:
$legend
[1] A=couche  B=lavage   C=peinture  D=tissu  E=protection #les 5 facteurs étudiés

$generators #Confusion avec les interactions du facteur 4 et 5
[1] D=AB E=AC


Alias structure: #générateurs d'alias
$main
[1] A=BD=CE B=AD    C=AE    D=AB    E=AC

$fi2
[1] BC=DE BE=CD

The design itself: #sortie des 8 essais à réaliser

   couche lavage  peinture     tissu protection
1 epaisse   doux acrylique polyester        non
2    fine   doux acrylique     coton        oui
3 epaisse   doux     posca polyester        oui
4    fine   doux     posca     coton        non
5    fine normal acrylique polyester        oui
6    fine normal     posca polyester        non
7 epaisse normal     posca     coton        oui
8 epaisse normal acrylique     coton        non

class=design, type= FrF2



#Qualité du plan

X<-model.matrix(~.,data=plan)
t(X)%*%X
            (Intercept) couche1 lavage1 peinture1 tissu1 protection1
(Intercept)           8       0       0         0      0           0
couche1               0       8       0         0      0           0
lavage1               0       0       8         0      0           0
peinture1             0       0       0         8      0           0
tissu1                0       0       0         0      8           0
protection1           0       0       0         0      0           8
```

Il également possible de le savoir par calcul. La résolution d'un plan correspond à la longueur du plus petit générateur d'alias. Le plan de résolution 2^{5-2} compte 3 générateurs d'alias (2^2-1=3).

La constante peut être confondue avec les interactions 123, 135 et 2345. Ainsi, ce plan est de résolution III. Les effets principaux sont confondus avec des interactions d'ordre 2 ou plus.

La sortie R permet de vérifier ces informations.

5. **Quelle suite imaginez-vous pour cette expérience ? Comment analyserez-vous les résultats d'expérience ?**

La suite logique serait d'effectuer tous ces essais dans la réalité. Une fois les t-shirts lavés, il serait intéressant d'interroger un maximum de personne et de leur demander de noter la tenue de la peinture sur une échelle de 0 (Le dessins est illisible, le peinture a été très dissoute au lavage) à 10 (le dessin est intact, la peinture a bien tenu) pour chacun des t-shirts.

Puis, il serait intéressant d'effectuer une analyse de variance à 6 facteurs (les 5 facteurs à 2 modalités étudiés + le facteur juge).

Exercice 8. Packaging le plus apprécié des consommateurs (exemple fictif, par Anaëlle Lecorgne)

Afin de choisir le visuel de l'emballage des nouveaux pavés végétaux qui sera le mieux perçu par les consommateurs, une analyse conjointe est réalisée. Pour cela, on demande à 150 participants de donner une note allant de 0 à 10 à différents packagings proposés. 6 critères sont étudiés :

- La couleur de fond : dégradé de blanc à vert ou de blanc à jaune
- La mention « végan » est indiquée ou non
- La mention « riche en protéines » est indiquée ou non
- Le nom du produit en haut ou à droite
- La photo : image des légumes à côté du pavé végétal ou non
- La DLC : sur le devant ou sur le côté de la boîte

Des packagings sont créés pour l'analyse conjointe, en associant toutes les combinaisons possibles. On limitera cependant le classement de 8 présentations emballages par participant.

1. **On décide de réaliser un plan en 8 essais. Donner le nom du plan, les générateurs d'alias et la résolution du plan, et construire le plan sur R.**

Plan en 8 essais : 2^{6-3}, qui correspond au plan de base 2^3. Il permet d'estimer les effets suivants :

I 1 2 3 12 13 23 123

On doit ajouter 3 facteurs au plan de base pour avoir les 6 effets. On confond 3 facteurs avec les interactions d'ordre le plus élevé -1, ici les interactions d'ordre 2.

On peut confondre le facteur 4 avec l'interaction 12, le facteur 5 avec l'interaction 13, le facteur 6 avec l'interaction 23.

On a les 3 générateurs d'alias initiaux suivants :

I_a= 124, I_b= 135, I_c= 236

On en déduit les autres générateurs d'alias en faisant les produits de ces générateurs initiaux entre eux, 2 à 2, 3 à 3 et le produit des 3. On obtient les générateurs d'alias suivants :

I_{ab}= 2345, I_{ac}= 1346, I_{bc}= 1256, I_{abc}= 456

On retrouve les 2^3-1= 7 générateurs d'alias attendus.

Le plan est de résolution III, puisque la plus petite interaction avec laquelle la constante est confondue est d'ordre 3. Pour estimer les effets principaux, il faut faire l'hypothèse que les interactions d'ordre 2 et plus sont négligeables.

Sur R :

```
library(FrF2)
plan<-
FrF2(nruns=8,nfactors=6,factor.names=list(couleur=c("vert","jaune"),vegan=c("oui","non"),protéines=c("oui","non"),nom=c("haut","droite"),photolegumes=c("oui","non"),DLC=c("devant","côté")))
summary(plan)
```

```
Call:
FrF2(nruns = 8, nfactors = 6, factor.names = list(couleur = c("vert",
    "jaune"), vegan = c("oui", "non"), protéines = c("oui",
    "non"), nom = c("haut", "droite"), photolegumes = c("oui",
    "non"), DLC = c("devant", "côté")))

Experimental design of type  FrF2
8  runs

Factor settings (scale ends):

Design generating information:
$legend
[1] A=couleur      B=vegan        C=protéines    D=nom
[5] E=photolegumes F=DLC

$generators
[1] D=AB E=AC F=BC

Alias structure:
$main
```

```
[1] A=BD=CE B=AD=CF C=AE=BF D=AB=EF E=AC=DF F=BC=DE

$fi2
[1] AF=BE=CD

The design itself:
class=design, type= FrF2
```

couleur <chr>	vegan <chr>	protéines <chr>	nom <chr>	photolegumes <chr>	DLC <chr>
vert	oui	oui	haut	oui	devant
jaune	non	non	droite	non	côté

2 rows

couleur <fctr>	vegan <fctr>	protéines <fctr>	nom <fctr>	photolegumes <fctr>	DLC <fctr>	
1	jaune	non	oui	droite	oui	devant
2	jaune	non	non	droite	non	côté
3	vert	oui	oui	droite	non	côté
4	jaune	oui	oui	haut	oui	côté
5	vert	non	oui	haut	non	devant
6	jaune	oui	non	haut	non	devant
7	vert	non	non	haut	oui	côté
8	vert	oui	non	droite	oui	devant

8 rows

```
X <- model.matrix(~.,data=plan)
t(X)%*%X
```

```
             (Intercept) couleur1 vegan1 protéines1 nom1 photolegumes1 DLC1
(Intercept)            8        0      0          0    0             0    0
couleur1               0        8      0          0    0             0    0
vegan1                 0        0      8          0    0             0    0
protéines1             0        0      0          8    0             0    0
nom1                   0        0      0          0    8             0    0
photolegumes1          0        0      0          0    0             8    0
DLC1                   0        0      0          0    0             0    8
```

2. Pourrait-on analyser l'interaction de la mention « riche en protéines » et mention « végan » ?

Ici, on s'intéresse à 6 facteurs, ainsi que la constante, ce qui demande l'estimation de 7 paramètres. Etant donné que l'on fait 8 essais, le plan n'est pas saturé, un dernier paramètre pourrait être estimé. Cependant, l'interaction des mentions « riche en protéines » et « végan » (interaction 23) est confondue avec un des effets principaux (le 6ème qui correspond à la DLC). Il ne pourra donc pas être analysé dans ce plan, il faudrait faire un nouveau plan en 16 essais.

3. Donner la méthode pour analyser ces résultats, ainsi que le modèle employé.

Pour analyser les résultats, on utilise l'analyse de variance à 8 facteurs dont 6 facteurs à deux modalités et le facteur juge à 150 modalités.

Le modèle étudié est donc le suivant :

$Y_{i,j,k,l,m,n,o}= \mu+\alpha_i+\beta_j+\gamma_k+\delta_l+\kappa_m+\nu_n+\tau_o+\xi_{ijklmno}$

Avec les hypothèses sur les résidus :

Pour tout (i,j,k,l,m,n,o), $L(\xi_{ijklmn,o}) = N(0, \sigma)$
Pour tout (i,j,k,l,m,n,o) ≠ (l',j',k',l',m',n',o'), $cov(\xi_{ijklmno},\xi_{i'j'k'l'm'n'o'}) = 0$

Exercice 9. Feu tricolore et fluidité du trafic (exemple fictif, par C. Fouilland)

Exercice issu d'un exemple fictif, inspiré d'études scientifiques.

L'objectif de cette étude est d'étudier le temps d'attente des véhicules au feu rouge. Réaliser des expériences sur le trafic routier est quasi-impossible puisque cela est coûteux, qu'il est difficile de reproduire exactement la même expérience plusieurs fois... C'est pour cela que les scientifiques ont mis au point des modèles hydrodynamiques pour modéliser le flux de véhicules. Le flux de voitures est assimilé à un flux laminaire de liquide se propageant dans un tube. Si le liquide est ralenti à un point donne du tuyau, le liquide s'accumulera en amont de ce point: un bouchon apparait. Il s'agit du même phénomène lorsque des véhicules doivent s'arrêter au feu rouge. On décide de mettre en place cette expérience et de s'intéresser aux facteurs suivants:

- A : la fluidité initiale de l'écoulement (la fluidité initiale du trafic) : fluide ou pas fluide
- B : les "remous" du liquide (la tendance des conducteurs à « sur-réagir ») : avec ou sans remous
- C : l'accumulation de liquide au point initial avant sa propagation en amont (la tendance des conducteurs à réagir en retard) : accumulation ou non
- D : la vitesse du liquide (la vitesse du véhicule) : rapide ou lente
- E : la densité du liquide (la densité du trafic) : densité élevée ou faible

Toutefois, la vitesse du fluide et sa densité ont globalement une faible incidence.

1. **Avec les moyens disponibles, on ne peut réaliser que 10 essais maximum. Quel type de plan de base allez-vous alors construire ? Justifier. Ecrire les intitulés des colonnes de la matrice des effets associée à ce plan (contenant tous les effets principaux et toutes les interactions).**

Il y a 5 facteurs et pour chacun de ces facteurs, les scientifiques se sont intéressés à 2 modalités : on va donc construire un plan fractionnaire. Le plan de base serait donc le plan 2^5 = 32 essais. Le nombres d'essais étant limités à 10, on choisit le plan fractionnaire 2^{5-2} = 8 essais.

I	A	B	C	AB	AC	BC	ABC

2. **A quelles interactions affectez les effets supplémentaires (facteurs D et E) ? Pourquoi ? Explicitez les générateurs d'alias et les autres confusions.**

On doit ajouter 2 facteurs supplémentaires, on va donc les confondre avec une interaction de l'ordre le plus élevé – 1. L'ordre le plus élevé étant 3, on confond D avec AB et E avec BC par exemple.

I	A	B	C	AB	AC	BC	ABC
ABD	BD	AD	ABCD	**D**	BCD	ACD	CD
BCE	ABCE	CE	BE	ACE	ABE	**E**	AE
ACDE	CDE	ABCDE	ADE	BCDE	DE	ABDE	BDE

D = AB => I = ABD

E= BC => I = BCE

I = ABD x BCE = ACDE

La première colonne correspond aux générateurs d'alias. Les autres colonnes correspondent aux confusions. On détail l'obtention des confusions de la 2e colonne (A), les autres colonnes sont obtenues de la même façon :

I=ABD => A(I) = A(ABD) => A=BD

I = BCE => A(I) = A(BCE) => A=ABCE

I = ACDE => A(I) = A(ACDE) => A = CDE

3. **Pouvez-vous estimer tous les effets ? Pourquoi ?**

Avec ces confusions telles quelles, on ne peut estimer que les colonnes dans leur ensemble (par paquet). Par exemple, on ne peut pas savoir ce qui est dû à l'interaction CDE ou au facteur A. Toutefois on sait que les interactions d'ordre supérieur à 3 sont négligeables. Pour estimer les effets principaux, on doit faire l'hypothèse que les interactions d'ordre 2 sont également négligeables. On voit ainsi qu'il est important de bien choisir les facteurs principaux et les facteurs supplémentaires.

4. **Quelle est la résolution de ce plan d'expérience ? Que faudrait-il faire si on ne pouvait pas considérer les interactions d'ordre 2 comme négligeables ?**

La résolution correspond à la longueur du plus petit générateur d'alias. Il s'agit donc d'un plan de résolution III. Les effets principaux sont donc confondus avec les interactions d'ordre 2 (ou plus). Si les interactions d'ordre 2 n'étaient pas négligeables, il faudrait augmenter la résolution et cela passe nécessairement par un nombre d'expériences plus élevé.

5. **Vous avez obtenu les paquets de I, A, B, C, D, AC, E et ABC. On associe à ces paquets des valeurs factuelles (présentées dans le tableau). On considère que les effets sont intéressants lorsque leur valeur absolue est plus grande que 1.**

I	A	B	C	AB	AC	BC	ABC
ABD	BD	AD	ABCD	**D**	BCD	ACD	CD
BCE	ABCE	CE	BE	ACE	ABE	**E**	AE
ACDE	CDE	ABCDE	ADE	BCDE	DE	ABDE	BDE
3,1	3,0	2,4	-1,8	4,1	0,64	-0,58	0,12

Déterminez les effets estimables et leur estimation, vous expliquerez votre démarche. Relevez une incohérence avec le sujet.

I	A	B	C	AB	AC	BC	ABC
ABD	BD	AD	ABCD	**D**	BCD	ACD	CD
BCE	ABCE	CE	BE	ACE	ABE	**E**	AE
ACDE	CDE	ABCDE	ADE	BCDE	DE	ABDE	BDE
3,1	3,0	2,4	-1,8	4,1	0,64	-0,58	0,12

Règle 1: Lorsqu'un paquet est négligeable (valeur absolue inférieure à 1), tous les termes du paquet sont négligeables: les termes des paquets AC, E et ABC sont donc négligeables.

Règle 5: Toutes les interactions d'ordre supérieures ou égales à deux sont négligeables.

Ainsi, on ne peut estimer uniquement les valeurs des effets A, B, C et D. L'effet du facteur A est égal à 3,0, celui du facteur B à 2,4, celui du facteur C à -1,8 et celui du facteur D à 4,1.

On observe que l'effet de la vitesse est le plus important. Or, d'après le sujet, la vitesse du fluide a globalement une faible incidence. Il y a donc une incohérence. Cela provient certainement du fait que l'on a négligé l'interaction AB en faveur du facteur D avec la règle 5. Or étant donné que les facteurs A et B ont une importance relativement élevée, on peut supposer que leur interaction peut aussi avoir une importance élevée. Il peut donc être intéressant de passer à la résolution 4 dans laquelle le facteur D ne sera plus confondu avec AB mais avec des interactions d'ordre 3 ou plus.

Exercice 10. Etude de l'hydratation des ingrédients du pain (exemple fictif, par Jean-Baptiste PETIT)

Les conditions sont fictives mais inspirées de données de stage.

On cherche à étudier le fonctionnement d'une nouvelle machine permettant d'hydrater certains ingrédients d'un pain complet aux céréales à très haute pression avant de les mélanger les uns aux autres. Le but est encore une fois d'augmenter la tendreté du pain de mie.

On étudie cette fois l'influence de l'hydratation (ou non) de certains ingrédients :

- Farine A
- Gluten B
- Céréales C
- Farine de soja D
- Son de blé E

1. **Donner le plan complet de tous les essais**

On étudie 5 paramètres à 2 modalités chacun (hydraté ou non), il s'agit donc d'un plan 2^5 avec 32 essais.

2. **On souhaite ne faire que 8 essais pour étudier ces paramètres. Donner ce plan avec le(s) générateur(s) d'alias et sa résolution. Que signifie cette résolution ?**

Il s'agit donc d'un plan 2^{5-2}. Pour réaliser ce plan, on prend au départ un plan à 3 facteurs, et on confond les 2 derniers avec les interactions d'ordre maximal – 1, ici d'ordre 2. Pour les générateurs d'alias, on a donc :

D = AB -> I = ABD
E = AC -> I = ACE
I = BCDE

Les plus petits générateurs d'alias sont les premiers, ce plan est de résolution 3.

On réalise ce plan sur R :

```
plan2 <-FrF2(nruns=8, nfactors=5, factor.names=list(Farine=c("hydraté","sec"), Gluten=c("hydraté","sec"), Céréales=c("hydraté","sec"), Soja=c("hydraté","sec"), Son=c("hydraté","sec")))

   Farine  Gluten Céréales    Soja     Son
1 hydraté hydraté      sec     sec hydraté
2     sec hydraté  hydraté hydraté hydraté
3     sec     sec  hydraté     sec hydraté
4 hydraté     sec  hydraté hydraté     sec
5     sec     sec      sec     sec     sec
6 hydraté hydraté  hydraté     sec     sec
7     sec hydraté      sec hydraté     sec
8 hydraté     sec      sec hydraté hydraté
class=design, type= FrF2
```

La résolution 3 signifie que les effets principaux sont confondus avec des interactions d'ordre 2, il faut donc considérer ces interactions comme négligeables pour pouvoir appliquer ce plan.

3. **On aimerait avoir un plan de résolution 4. En quoi un tel plan est-il plus précis ? Donner ce plan en précisant bien le nombre d'essais à réaliser.**

Ce plan est plus précis car on peut confondre les effets principaux avec des interactions d'ordre 3, qui sont plus simples à négliger.

```
plan4 <-FrF2(nfactors=5, resolution=4, factor.names=list(Farine=c("hydraté","sec"),
Gluten=c("hydraté","sec"), Céréales=c("hydraté","sec"), Soja=c("hydraté","sec"),
Son=c("hydraté","sec")))
    Farine  Gluten Céréales    Soja     Son
1  hydraté hydraté      sec     sec     sec
2  hydraté     sec  hydraté hydraté hydraté
3  hydraté hydraté      sec hydraté hydraté
4  hydraté hydraté  hydraté     sec hydraté
5      sec hydraté  hydraté     sec     sec
6      sec     sec  hydraté     sec hydraté
7  hydraté hydraté  hydraté hydraté     sec
8  hydraté     sec      sec     sec hydraté
9      sec     sec  hydraté hydraté     sec
10     sec hydraté      sec     sec hydraté
11     sec     sec      sec hydraté hydraté
12 hydraté     sec      sec hydraté     sec
13     sec     sec      sec     sec     sec
14     sec hydraté      sec hydraté     sec
15 hydraté     sec  hydraté     sec     sec
16     sec hydraté  hydraté hydraté hydraté
class=design, type= FrF2
```

Pour obtenir un plan d'une telle qualité, il faut donc réaliser 16 essais au lieu de 8.

Exercice 11. Plan d'expérience pour optimiser la production d'algues (exemple réel, par Andrialovanirina Nicolas)

Dans le cadre de la situation d'anti-gaspillage, un producteur d'algue marine veut optimiser (éviter le plus le gaspillage) sa production (sans pour autant diminuer sa production). Pour ce faire, il a décidé de réaliser 1 expérience sur ces 8 bassins sur 5 paramètres différents pendant un mois de production (l'algue se multiple à environ 100 fois environ durant cette période).

Il voudrait voir quels paramètres qu'il peut maîtriser influence sa production, et quel paramètre nécessitant de l'énergie peut-il délaisser.

Chaque paramètre a 2 modalités :

- température de l'eau : (1) non stabilisé dépendant de la température ambiante ou (2) stabilisée à 35 °C
- brassage de l'eau : (1) naturel par le vent ou (2) par pompe à brassage
- source de lumière : (1) solaire ou (2) par des lampes
- nitrate de potassium (kno3) pour le milieu de culture : (1) 0,5 kg ou (2) 1 kg
- alimentation : (1) 0,5 litre d'urine + 100ml de solution de Fer ou (2) 1 litre d'urine + 200ml de solution de Fer

Vous êtes en charge de réaliser les expériences pour une durée de 30 jours, vous devriez réaliser un plan d'expérience sur ces 8 bassins.

1. **Le producteur vous demande combien d'essais doit-on fait pour tester tous les paramètres ?**

Il faudra au total faire $2^5 = 32$ essais.

2. **Pour ce 8 bassins donc, quel plan devriez-vous réaliser pour ces 5 paramètres ?**

Vu que l'on a 2 modalités pour chaque paramètre, on va utiliser un plan fractionnaire à $2^{5-2} = 8$ essais.

3. **Faites la construction du plan sur R (dites-nous toutes les confusions et la résolution du plan).**

```
library(FrF2)
set.seed(123)
plan.ex1 <- FrF2(nruns = 8, nfactors = 5,
                 factor.names = list(
                   temperature = c("non_stabiliser", "stabiliser"),
                   brassage = c("non", "oui"),
                   lumiere = c("solaire", "lampe"),
                   kno3 = c("0,5", "1"),
                   uree = c("0,5", "1")))
summary(plan.ex1)

# Le plan construit
Call:
FrF2(nruns = 8, nfactors = 5, factor.names = list(temperature = c("non_stabiliser",
    "stabiliser"), brassage = c("non", "oui"), lumiere = c("solaire",
    "lampe"), kno3 = c("0,5", "1"), uree = c("0,5", "1")))

Experimental design of type  FrF2
8  runs #nombre d'essais

# Les facteurs avec chaque modalité
Factor settings (scale ends):
     temperature brassage    lumiere kno3  uree
1 non_stabiliser non      solaire 0,5  0,5
2 stabiliser     oui      lampe   1    1

Design generating information:
$legend
[1] A=temperature B=brassage    C=lumiere     D=kno3        E=uree

#générateur d'alias de D et E (confondus avec les interactions les plus élevé -1)
$generators
[1] D=AB  E=AC

Alias structure:
$main
[1] A=BD=CE  B=AD     C=AE     D=AB     E=AC

#Confusion entre les interactions d'ordre 2
$fi2
[1] BC=DE  BE=CD

#Confusion des effets
```

I	A	B	C	AB	AC	BC	CD	ABC
ABD	BD	AD	ABCD	D	BCD	ACD	ABC	CD
ACE	CE	ABCE	AE	BCE	E	ABE	ADE	BE
BCDE	ABCDE	CDE	BDE	ACDE	ABDE	DE	BE	ADE

Le plan est de résolution III car la constante (I) est confondue avec des interactions d'ordre 3 (interaction d'ordre le moins élevé). Et les effets principaux sont confondus avec des interactions d'ordre 2 ou plus.

```
#Les 8 expériences à réaliser
The design itself:
```

```
    temperature brassage lumiere kno3 uree
1 non_stabiliser      oui   lampe  0,5  0,5
2     stabiliser      oui   lampe    1    1
3 non_stabiliser      oui solaire  0,5    1
4     stabiliser      non   lampe  0,5    1
5     stabiliser      non solaire  0,5  0,5
6     stabiliser      oui solaire    1  0,5
7 non_stabiliser      non   lampe    1  0,5
8 non_stabiliser      non solaire    1    1
class=design, type= FrF2
```

Vu qu'ici on a qu'un mois d'expérience pour les 8 bassins (donc seulement 8 essais), on ne peut pas faire plus d'essais et donc on ne peut pas étudier les interactions. De ce fait, on ne va étudier que les 5 facteurs (les effets principaux).

4. **Vérifier la qualité du plan d'expérience**

```
options(contrasts=c("contr.sum", "contr.sum"))
X <- model.matrix(~., data = plan.ex1)
solve(t(X)%*%X)
             (Intercept) temperature1 brassage1  lumiere1    kno31  uree1
(Intercept)        0.125        0.000     0.000     0.000    0.000  0.000
temperature1       0.000        0.125     0.000     0.000    0.000  0.000
brassage1          0.000        0.000     0.125     0.000    0.000  0.000
lumiere1           0.000        0.000     0.000     0.125    0.000  0.000
kno31              0.000        0.000     0.000     0.000    0.125  0.000
uree1              0.000        0.000     0.000     0.000    0.000  0.125
```

On a un plan orthogonal où les estimations de chaque facteur sont indépendantes l'une de l'autre. On minimise la variance Var(β) avec une valeur de $(X'X)^{-1}$ assez faible (avec Var(β) = $(X'X)^{-1}\sigma^2$).

Exercice 12. Maximiser le rendement d'une prairie en fonction des conditions de culture (contexte réel, par AV. de Croutte)

Nous nous trouvons dans un laboratoire de recherche sur les pâtures en Nouvelle-Zélande. Nous étudions ici les capacités de rendement d'un mix de 2 espèces : Ray-grass anglais et Trèfle blanc. Nous voulons étudier quelles conditions de cultures donneront un meilleur rendement. Nous pouvons jouer sur 4 paramètres : l'irrigation, la méthode de semis, la méthode de coupe (1 fois par mois) et la quantité de fertilisation azotée.

Nous avons donc 3 facteurs à 2 modalités et 1 facteur à 4 modalités :

- Irrigation : avec ou sans (+1 ou -1),
- Méthode de semis : rangs alternés ou mélangés (+1 ou -1),
- Méthode de coupe : moutons ou tondeuse (+1 ou -1),
- Fertilisation azotée : 0 ou 100 ou 200 ou 300 kg N/ha (1, 2, 3 ou 4).

Cependant la surface dont nous disposons pour monter l'expérience ne peut contenir que 8 parcelles. Il faut donc construire un plan en 8 essais pour étudier simultanément ces 3 facteurs.

Dans un premier temps, on considèrera le facteur 'fertilisation azotée' à 4 modalités comme 2 facteurs 'azote1' et 'azote2' chacun à 2 modalités '1' et '2'.

1. **Quel serait le plan complet et combien d'essais nécessiterait-il ? Avec ces 8 essais, quel est le nom du plan à construire ?**

On a 5 facteurs à 2 modalités à tester, donc le plan complet serait 2^5, donc en 32 essais. Mais on se réduit à 8 essais, on devra donc construire le plan $L^8 2^{5-2}$, ce qui engendrera des confusions.

2. **Décrire la méthode de construction du plan et le construire sur R.**

On doit choisir 3 facteurs principaux (plan 23) et confondre les 2 qu'il reste avec des interactions d'ordre 2. On peut donc ici confondre azote1 avec l'interaction irrigation-méthode de semis, et azote2 avec l'interaction irrigation-méthode de coupe.

On construit le plan sur R avec la fonction FrF2 du package FrF2 :

```
library(FrF2)
plan <- FrF2(nruns=8,nfactors=5,factor.names=list(irrigation=c("oui","non"),
semis=c("alternés","mélangés"),coupe=c("moutons","tondeuse"),
```

```
azote1=c("1","2"),azote2=c("1","2")))
summary(plan)

Call:
FrF2(nruns = 8, nfactors = 5, factor.names = list(irrigation = c("oui",
    "non"), semis = c("alternés", "mélangés"),
    coupe = c("moutons", "tondeuse"), azote1 = c("1",
        "2"), azote2 = c("1", "2")))

Experimental design of type  FrF2
8  runs

Factor settings (scale ends):
  irrigation    semis    coupe azote1 azote2
1        oui alternés  moutons      1      1
2        non mélangés tondeuse      2      2

Design generating information:
$legend
[1] A=irrigation B=semis      C=coupe      D=azote1     E=azote2

$generators
[1] D=AB E=AC


Alias structure:
$main
[1] A=BD=CE B=AD     C=AE     D=AB     E=AC

$fi2
[1] BC=DE BE=CD


The design itself:
  irrigation    semis    coupe azote1 azote2
1        non mélangés tondeuse      2      2
2        non alternés tondeuse      1      2
3        oui alternés  moutons      2      2
4        oui mélangés  moutons      1      2
5        non mélangés  moutons      2      1
6        non alternés  moutons      1      1
7        oui alternés tondeuse      2      1
8        oui mélangés tondeuse      1      1
class=design, type= FrF2
```

3. Donner les générateurs d'alias et la résolution du plan.

On ajout 2 facteurs au plan, donc on doit obtenir 3 générateurs (2^2-1).

En reprenant les notations de R, on a les confusion D=AB et E=AC, donc ABD et ACE sont des générateurs d'alias. On le multiplie entre eux et on trouve le 3e générateur ABDACE = BCDE.

Donc I = ABD = ACE = BCDE.

La plus petite interaction confondue avec la constante I est d'ordre 3, donc le plan est de résolution III. Cela signifie qu'on ne peut analyser sans ambigüité que les effets principaux.

4. Comment faire pour récupérer notre facteur azote à 4 modalités ? Donner la matrice des essais.

On va utiliser les 2 facteurs à 2 modalités azote1 et azote2 pour construire le facteur azote à 4 modalités. On fera donc la construction d'un plan $L_8 2^3 4$. On définit les équivalences suivantes :

azote1	**azote2**	**azote**
1	1	0
1	2	100
2	1	200
2	2	300

On peut récupérer la matrice des essais précédemment calculée dans R et ajouter ce facteur azote. On considère les modalités binaires comme +1 et -1, et celles du facteur azote comme 1, 2, 3 et 4. On a donc la matrice des essais suivante :

	Irrigation	**Semis**	**Coupe**	**Azote**
1er essai	-1	-1	-1	4

2e essai	-1	1	-1	2
3e essai	1	1	1	4
4e essai	1	-1	1	2
5e essai	-1	-1	1	3
6e essai	-1	1	1	1
7e essai	1	1	-1	3
8e essai	1	-1	-1	1

5. Commenter la qualité du plan.

On peut construire la matrice des effets X, puis la matrice de dispersion $(X'X)^{-1}$.

$$X = \left(\rightarrow X = \begin{array}{ccccccc} 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & -1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 1 \\ 1 & -1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & -1 & 0 & 0 & 1 \\ 1 & 1 & -1 & -1 & 1 & 0 & 0 \end{array} \right) \begin{array}{c} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{array} \quad \left(\begin{array}{cccccc} 1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 0 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & 0 & 1 \\ 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & -1 & 0 & 0 \\ 1 & 1 & -1 & -1 & 1 & 0 \end{array} \right) \begin{array}{|c} -1 \\ 0 \\ -1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{array}$$

```
X <- matrix(c(1,1,1,1,1,1,1,1,-1,-1,1,1,-1,-1,1,1,-1,1,1,-1,-1,1,1,-1,-1,
-1,1,1,1,1,-1,-1,-1,0,-1,0,0,1,0,1,-1,1,-1,1,0,0,0,0,-1,0,-1,0,1,0,1,0),nrow=8)
X
     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    1   -1   -1   -1   -1   -1   -1
[2,]    1   -1    1   -1    0    1    0
[3,]    1    1    1    1   -1   -1   -1
[4,]    1    1   -1    1    0    1    0
[5,]    1   -1   -1    1    0    0    1
[6,]    1   -1    1    1    1    0    0
[7,]    1    1    1   -1    0    0    1
[8,]    1    1   -1   -1    1    0    0

solve(t(X)%*%X)
      [,1]  [,2]  [,3]  [,4]   [,5]   [,6]   [,7]
[1,] 0.125 0.000 0.000 0.000  0.000  0.000  0.000
[2,] 0.000 0.125 0.000 0.000  0.000  0.000  0.000
[3,] 0.000 0.000 0.125 0.000  0.000  0.000  0.000
[4,] 0.000 0.000 0.000 0.125  0.000  0.000  0.000
[5,] 0.000 0.000 0.000 0.000  0.375 -0.125 -0.125
[6,] 0.000 0.000 0.000 0.000 -0.125  0.375 -0.125
[7,] 0.000 0.000 0.000 0.000 -0.125 -0.125  0.375
```

On a bien tous les facteurs orthogonaux, même au facteur 4 (l'azote, décrit par les colonnes et lignes 5 à 7), et des valeurs de variances faibles, donc le plan est de bonne qualité.

Exercice 13. Plan fractionnaire et application à l'agronomie (exemple fictif, par E. David)

L'exercice suivant se base sur une thématique et des données fictives. Un agriculteur souhaite introduire des cultures en association sur son parcellaire. Ils souhaitent tester la meilleure association entre 4 espèces végétales : 2 espèces de blé et 2 espèces de pois. Il étudie également l'effet de deux différentes dates de semis. Pour évaluer l'association, ils étudient le rendement total (rendement en pois + rendement en blé).

- A : Blé dur issu de la sélection variétale (2 modalités : Anvergur et Nobilis)

- B : Blé ancestrale (2 modalités : Touselles de Mayan et Bladette de Puylaurens)
- C : Pois protéagineux d'Hiver issu de la sélection variétale (2 modalités : Lapony et Escrime)
- D : Pois protéagineux d'Hiver ancestrale (2 modalités : Fresnel et Aviron)
- E : Date de semis (2 modalités : 15 octobre et 15 novembre)

On étudie alors l'influence de ces cinq facteurs qualitatifs qu'on note A, B, C, D, E présentant chacun 2 modalités qu'on notera 1 et 2 sur la réponse Y du rendement. Au vu des contraintes budgétaires de l'agriculteur pour tester les associations, il souhaite construire un plan permettant d'étudier 5 facteurs à 2 niveaux en 16 essais.

1. **Définir le plan de base à choisir pour étudier 5 facteurs en 16 essais**

Pour étudier un plan complet à 5 facteurs à 2 niveaux, il faut $2^5 = 32$ essais.
Le plan de base est un plan complet en 16 essais, c'est le plan 2^4. La première ligne de la matrice des effets est la suivante :

I	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD

2. **Déterminer l'effet confondu avec le facteur E et déduire les autres confusions du plan. Indiquer la résolution du plan.**

On ajoute un facteur donc on affecte la confusion avec l'interaction d'ordre le plus élevé. On a donc :
E = ABCD => EE = ABCDE => I = ABCDE. E est confondu avec ABCD.
Le seul générateur d'alias est I = ABCDE. Cette constante I est confondue avec l'interaction ABCDE donc il découle les confusions suivantes :

I	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
ABCDE	BCDE	ACDE	ABDE	ABCE	CDE	BDE	BCE	ADE	ACE	ABE	DE	CE	BE	AE	E

Dans ce plan, les facteurs principaux A, B, C, D sont confondus avec des interactions d'ordre 4 ou plus et le générateur d'alias est confondu avec une interaction d'ordre 5. Le plan est donc de résolution V. Ceci implique que les effets principaux sont confondus avec les interactions d'ordre 4 ou plus et les interactions d'ordre 2 sont confondues avec les interactions d'ordre 3 ou plus. Les effets principaux sont estimables si l'on considère que les interactions d'ordre 4 ou plus sont négligeables. Dans ce cas, l'agriculteur néglige les interactions d'ordre 4 et ne s'intéresse qu'aux interactions d'ordre 2 pour étudier les échanges trophiques entre les légumineuses (pois) et le blé.

3. **L'agriculteur souhaite tester les interactions entre les espèces introduites, il s'agit des interactions d'ordre 2. Est-ce possible avec le plan construit ?**

À partir des données, on peut estimer : 1 paramètre pour la constante, 5 paramètres pour chacun des facteurs. Comme on peut réaliser 16 essais au total, il reste (16 – 6) soit 10 degrés de liberté disponibles. On va donc pouvoir estimer sans ambiguïté 10 paquets d'effets contenant des interactions d'ordre 2 intéressant l'agriculteur.

4. **Construire le plan sous R. Calculer la matrice $(X'X)^{-1}$ associée au modèle avec seulement les effets principaux puis celle associée au modèle avec les interactions d'ordre 2.**

```
# Construction du plan de base en 16 essais : plan 24
library(FrF2)
plan = FrF2(nruns=16, nfactors=5, factor.names=list(A=c("A1","A2"), B=c("B1","B2"),
C=c("C1","C2"), D=c("D1","D2"), E=c("E1","E2")))
summary(plan)

Call:
FrF2(nruns = 16, nfactors = 5, factor.names = list(A = c("A1",
    "A2"), B = c("B1", "B2"), C = c("C1",
    "C2"), D = c("D1", "D2"), E = c("E1",
    "E2")))

Experimental design of type  FrF2
16  runs

Factor settings (scale ends):
   A  B  C  D  E
1 A1 B1 C1 D1 E1
2 A2 B2 C2 D2 E2

Design generating information:
$legend
[1] A=A B=B C=C D=D E=E

$generators
[1] E=ABCD

Alias structure:
[[1]]
```

```
[1] no aliasing among main effects and 2fis

The design itself:
    A  B  C  D  E
1  A1 B1 C1 D2 E1
2  A1 B1 C2 D2 E2
3  A1 B2 C1 D1 E1
4  A2 B2 C2 D2 E2
5  A2 B1 C1 D1 E1
6  A2 B2 C2 D1 E1
7  A2 B1 C1 D2 E2
8  A2 B1 C2 D1 E2
9  A1 B1 C1 D1 E2
10 A1 B1 C2 D1 E1
11 A1 B2 C1 D2 E2
12 A2 B2 C1 D1 E2
13 A2 B2 C1 D2 E1
14 A2 B1 C2 D2 E1
15 A1 B2 C2 D1 E2
16 A1 B2 C2 D2 E1
class=design, type= FrF2
```

```
# Construction de (X'X)-1 avec les effets principaux sans interaction
options(contrasts=c("contr.sum","contr.sum"))
X = model.matrix(~.,data=plan)
t(X)%*%X
solve(t(X)%*%X)
```

```
            (Intercept)     A1     B1     C1     D1     E1
(Intercept)      0.0625 0.0000 0.0000 0.0000 0.0000 0.0000
A1               0.0000 0.0625 0.0000 0.0000 0.0000 0.0000
B1               0.0000 0.0000 0.0625 0.0000 0.0000 0.0000
C1               0.0000 0.0000 0.0000 0.0625 0.0000 0.0000
D1               0.0000 0.0000 0.0000 0.0000 0.0625 0.0000
E1               0.0000 0.0000 0.0000 0.0000 0.0000 0.0625
```

Les effets principaux ne sont pas confondus entre eux.

```
# Construction de (X'X)-1 avec les interactions d'ordre 2
X = model.matrix(~A+B+C+D+E+A:B+A:C+A:D+A:B+C:B+D:C+D, data=plan)
solve(t(X)%*%X)
```

```
           (Intercept)     A1     B1     C1     D1     E1  A1:B1  A1:C1  A1:D1  B1:C1
(Intercept)      0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
A1               0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
B1               0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
C1               0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
D1               0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000
E1               0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000
A1:B1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000
A1:C1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000
A1:D1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000
B1:C1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625
C1:D1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
B1:D1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
             C1:D1  B1:D1
(Intercept) 0.0000 0.0000
A1          0.0000 0.0000
B1          0.0000 0.0000
C1          0.0000 0.0000
D1          0.0000 0.0000
E1          0.0000 0.0000
A1:B1       0.0000 0.0000
A1:C1       0.0000 0.0000
A1:D1       0.0000 0.0000
B1:C1       0.0000 0.0000
C1:D1       0.0625 0.0000
B1:D1       0.0000 0.0625
```

Les interactions d'ordre 2 ne sont pas confondues entre elles car il n'y aucune colonne égale. Un plan en 16 essais permet donc d'estimer les 5 interactions d'ordre 2 comme nous l'avions dit dans la question 3. L'agriculteur pourra avoir une idée de l'effet d'interaction entre les variétés de pois entre elles, entre les variétés de blé entre elles et également les interactions entre les variétés de pois et de blé.

Exercice 14. Comment détrôner le Coca Cola ? (exemple fictif, par Tom AYRAULT)

Une entreprise prête à tout pour détrôner le Coca Cola tente de mettre au point un nouveau soda mais peine à trouver une recette parfaite. Elle hésite sur quelques points :

- L'acidité (ajout fort ou faible d'acide citrique) ➔ facteur F1 à 2 modalités
- Le caractère pétillant (peu pétillant ou très pétillant) ➔ facteur F2 à 2 modalités
- Le goût (cerise ou pomme) ➔ facteur F2 à 2 modalités
- La quantité de sucre (peu sucré ou sucré) ➔ facteur F4 à 2 modalités

On veut faire goûter les différentes recettes à un panel. Chaque testeur devra donner une note entre 0 et 10 mais devra cependant ne tester que huit échantillons car on considère qu'au-delà de 8 échantillons, il est difficile de bien noter.
L'objectif est de construire un plan d'expérience.

1. Combien d'essais sont nécessaires pour construire un plan complet à 4 facteurs ?

Pour étudier un plan complet à 4 facteurs à deux modalités, il faut $2^4 = 16$ essais.

2. Pour quel plan de base doit-on opter ? Dessiner la matrice des effets du modèle saturé.

On a droit à 8 essais. Il y a 4 facteurs à deux modalités à étudier.
On a donc un plan $2^{4-1} = 8$ essais. Le plan de base est ainsi le plan 2^3.
On obtient la matrice des effets suivantes :

I	1 = F1	2 = F2	3 = F3	12	13	23	123
1	1	1	1	1	1	1	1
1	1	1	-1	1	-1	-1	-1
1	1	-1	1	-1	1	-1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	1	-1	-1	1	-1
1	-1	1	-1	-1	1	-1	1
1	-1	-1	1	1	-1	-1	1
1	-1	-1	-1	1	1	1	-1

3. Avec quel effet choisissez-vous de confondre le facteur 4 ? Justifier.

On choisit de confondre le facteur 4 avec l'interactions des effets 123. En effet lorsqu'on ajoute un facteur lors du choix des confusions, la confusion s'effectue avec l'interaction d'ordre le plus élevé, donc celle qui a le plus de chances d'être négligeable. En effet en supposant que les effets d'ordre 3 ou plus sont négligeables, on peut alors bien estimer F4.

4. Déduire alors toutes les autres confusions du plan ainsi que la résolution de ce plan.

En confondant le facteur 4 avec l'interaction 123 (4 = 123), on obtient que 44 = 1234 ➔ I = 1234.
On en déduit aussi que 1 = 234, 2 = 134, 3 = 124.
On a alors un plan de résolution IV (car I est confondu avec au minimum une interaction d'ordre 4).

5. Construire le plan sur R et vérifier l'exactitude des générateurs d'alias et la résolution du plan.

```
library(FrF2)
plan=FrF2(nruns = 8, nfactors = 4, factor.names = list(acidite=c("faible","forte"), petillant=c(
"peu", "très"), gout=c("cerise","pomme"), sucre=c("peu sucre", "sucre")), randomize=FALSE)
summary(plan)
```

[…]

```
Design generating information:
$legend
[1] A=acidite    B=petillant C=gout      D=sucre

$generators
[1] D=ABC    # On a bien I = ABCD (résolution IV)


Alias structure:
$fi2
[1] AB=CD AC=BD AD=BC    # On retrouve les confusions

[…]
```

L'entreprise se rend finalement compte que le facteur sucre est le plus important et qu'il nécessite une découpe plus fine en quatre modalités pour avoir une vue plus claire. Il faut donc revoir en urgence le plan d'expérience.

6. En repartant de la matrice des effets, expliquer avec quel(s) effet(s) choisissez-vous de confondre le facteur 4.

NB : Un facteur à quatre modalités équivaut à deux facteurs à deux modalités, ainsi on aura - - = 1, - + = 2, + - = 3, + + = 4.

On peut considérer le facteur F4 à quatre modalités comme deux facteurs (A et B) à deux modalités. On choisit alors de confondre A avec 12 et B avec 13. En effet pour l'ajout de deux facteurs ou plus, les confusions se font avec les interactions d'ordre le plus élevé **-1**.

7. En déduire la matrice des essais associée

	F1	F2	F3	F4
1er essai	1	1	1	4
2ème essai	1	1	-1	3
3ème essai	1	-1	1	2
4ème essai	1	-1	-1	1
5ème essai	-1	1	1	1
6ème essai	-1	1	-1	2
7ème essai	-1	-1	1	3
8ème essai	-1	-1	-1	4

Une fois le plan d'expérience achevé, on fait alors appel à un panel de 100 personnes pour goûter différentes versions de la boisson pour déterminer le mélange parfait.

8. Quel modèle utiliser pour analyser ces résultats ?

On peut effectuer une analyse de variance à 5 facteurs (3 facteurs à deux modalités (F1, F2, F3), 1 facteur à 4 modalités (F4) et 1 facteur à 100 modalités (facteur juge)).

Exercice 15. Masques pour le déconfinement (exemple fictif, par Junyi Zhao)

Au déconfinement le 11 mai 2020, un grand magasin parisien veut équiper ses collaborateurs de masques. Il souhaite créer un masque qui permet une utilisation prolongée (maximum 4 h) avec un confort optimal.

Il a le choix entre plusieurs variables :

- Couleur du masque (bleu, blanc)
- Matière (tissu, papier)
- Couture des élastiques (sur les oreilles ou derrière la tête)
- Forme (bec de canard ou rectangulaire)
- Produit en France (mention « fabriqué en France » présente ou non)
- Réutilisable (lavable ou non)

Il est décidé de faire 8 prototypes de masques.

1. Combien d'essais sont nécessaires pour construire le plan complet ?

Le plan complet comporte $2^6 = 64$ essais.

2. Quel est le plan à construire pour étudier les 6 facteurs en 8 essais ?

Le plan de base à construire est le plan $2^{6-3} = 2^3$.

3. Construire ce plan sous R avec la fonction *FrF2* du package *FrF2.* Quelle est la résolution de ce plan ?

On utilise le plan de base 2^3 qui est le plan en 8 essais

I 1 2 3 12 13 23 123

On doit confondre 3 facteurs avec 3 interactions :

Le facteur 4 avec l'interaction 12, le facteur 5 avec 13 et le facteur 6 avec 23.

C'est un plan de résolution 3 : la plus petite confustion avec la constante est d'ordre 3.

```
library(FrF2)
plan <- FrF2(nruns=8,nfactors=6,
             factor.names=list(couleur=c("bleu","blanc"), matiere=c("tissu","papier"),
                          couture=c("dessus","cote"), forme=c("canard","rectangle"),
                          france=c("oui","non"), lavable=c("oui","non")))
summary(plan)
```

```
Call:
FrF2(nruns = 8, nfactors = 6, factor.names = list(
couleur = c("bleu","blanc"),
matiere = c("tissu", "papier"),
couture = c("dessus", "cote"),
forme = c("canard","rectangle"),
france = c("oui", "non"),
lavable = c("oui", "non")))


Experimental design of type  FrF2
8  runs

Factor settings (scale ends):
  couleur matiere couture     forme france lavable
1    bleu   tissu  dessus    canard    oui     oui
2   blanc  papier    cote rectangle    non     non

Design generating information:
$legend
[1] A=couleur B=matiere C=couture D=forme   E=france  F=lavable

$generators
[1] D=AB E=AC F=BC


Alias structure:
$main
[1] A=BD=CE B=AD=CF C=AE=BF D=AB=EF E=AC=DF F=BC=DE

$fi2
[1] AF=BE=CD


The design itself:
  couleur matiere couture     forme france lavable
1   blanc   tissu  dessus    canard    oui     non
2    bleu   tissu  dessus rectangle    non     non
3    bleu   tissu    cote rectangle    oui     oui
4    bleu  papier  dessus    canard    non     oui
5    bleu  papier    cote    canard    oui     non
6   blanc   tissu    cote    canard    non     oui
7   blanc  papier  dessus rectangle    oui     oui
8   blanc  papier    cote rectangle    non     non
class=design, type= FrF2
```

Exercices sur les plans continus et surface de réponses

Exercice 1. Production de chocolats (exemple réel, par Hector Chassagnon)

Intro : Contexte

Un nouveau producteur de chocolat désire réaliser une nouvelle gamme de tablette de chocolat. Pour ça il a le choix entre des fèves récoltées entre 2006 et 2016. De plus pour sa nouvelle tablette, il hésite sur le pourcentage de cacao à mettre (entre 65% et 85%).

Pour être compétitif sur le marché il souhaite avoir le meilleur chocolat possible. Il cherche donc la meilleure combinaison possible pour maximiser l'appréciation de son chocolat. Il fait appel à vous pour le conseiller.

Partie 1

4. **Quelle serait votre approche face à cette expérience et votre stratégie de planification ?**

On cherche à maximiser la note d'appréciation du chocolat notée sur 5.

La note d'appréciation comprise entre 0 et 5 est la variable à expliquer. On cherche à l'expliquer par deux variables quantitatives, le pourcentage de cacao (compris entre 65% et 85%) et l'année de récolte (entre 2006 et 2014).

Pour trouver la combinaison qui optimise Y, on réalisera un plan d'expérience où nous testerons plusieurs combinaisons d'année de récolte et de pourcentage de cacao. Nous jugerons la qualité du plan, puis nous essaierons de modéliser l'appréciation de Y selon ces deux variables.

5. **Expliciter le modèle que vous utiliserez pour analyser les résultats d'un tel plan.**

Pour analyser les résultats d'un tel plan, on utilise classiquement le modèle de régression linéaire ou :

La note Y est expliquée par les deux variables quantitatives (β1x1, β2x2), leurs effets quadratiques et leurs interactions entre eux. On peut modéliser ce modèle par l'équation suivante :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon$$

6. **Quel plan mettez-vous en place ? Combien d'essais sont nécessaires ?**

Nous réalisons un plan composite centré à 2 facteurs. Comme mon nombre de facteur est faible (2) je prends le plan factoriel complet (2^2=4essais). J'ai besoin de 2*k, soit 4 essais, pour mes points en étoiles. Et comme je veux l'orthogonalité des facteurs je fais 8 essais au centre. Soit au total 16 essais.

7. **A l'aide de R, donnez la matrice des essais.**

```
library(rsm)
plan1<-ccd(2,coding = list(x1~(cacao-75)/20,x2~(AnneeRecolte-2011)/5), randomize = FALSE)
plan1<-plan1[,3:4]
round(plan1,0)
   cacao AnneeRecolte
1     55         2006
2     95         2006
3     55         2016
4     95         2016
5     75         2011
6     75         2011
7     75         2011
8     75         2011
9     55         2011
10    95         2011
11    75         2006
12    75         2016
13    75         2011
14    75         2011
15    75         2011
16    75         2011
```

8. **Analyser le plan.**

```
#Qualité du plan
X <- model.matrix(~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data=plan1)
```

```
XX<-t(X)%*%X
XX
              (Intercept) x1 x2 I(x1^2) I(x2^2) I(x1 * x2)
(Intercept)            16  0  0       8       8          0
x1                      0  8  0       0       0          0
x2                      0  0  8       0       0          0
I(x1^2)                 8  0  0      12       4          0
I(x2^2)                 8  0  0       4      12          0
I(x1 * x2)              0  0  0       0       0          4
round(solve(XX),3)
              (Intercept)    x1    x2 I(x1^2) I(x2^2) I(x1 * x2)
(Intercept)         0.125 0.000 0.000  -0.062  -0.062       0.00
x1                  0.000 0.125 0.000   0.000   0.000       0.00
x2                  0.000 0.000 0.125   0.000   0.000       0.00
I(x1^2)            -0.062 0.000 0.000   0.125   0.000       0.00
I(x2^2)            -0.062 0.000 0.000   0.000   0.125       0.00
I(x1 * x2)          0.000 0.000 0.000   0.000   0.000       0.25
```

Grace a ces deux matrices on peut voir qu'il y a une bonne orthogonalité entre les facteurs testés. Ce qui signifie que le plan d'expérience réalisée est bon. Il y a peu de confusion entre les variables, leurs effets quadratiques et leurs interactions. Les effets quadratiques sont très légèrement confondus avec la constant, mais les coefficients sont faibles donc on peut considérer la confusion comme négligeable.

Partie 2 : Résultats

Les résultats des expériences ont permis d'obtenir ces notes

```
Y<-c(3.2,4.4,2.3,3.9,3.4,3.3,3.7,3.5,2.7,4.2,3.8,3.4,3.3,3.6,2.9,3.8)
plan<-cbind.data.frame(plan1,Y)
```

9. **Analyser les résultats obtenus après la réalisation des essais. Peut-on faire confiance au modèle ?**

```
CR.rsm<-rsm(Y~FO(x1,x2)+TWI(x1,x2)+PQ(x1,x2), data=plan1)
summary (CR.rsm)
Call:
rsm(formula = Y ~ FO(x1, x2) + TWI(x1, x2) + PQ(x1, x2), data = plan1)

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  3.437500   0.094511 36.3714 5.868e-12 ***
x1           0.615165   0.094511  6.5089 6.818e-05 ***
x2          -0.245711   0.094511 -2.5998    0.0265 *
x1:x2        0.100000   0.133659  0.7482    0.4716
x1^2        -0.012500   0.094511 -0.1323    0.8974
x2^2         0.062500   0.094511  0.6613    0.5234
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.8337,     Adjusted R-squared:  0.7506
F-statistic: 10.03 on 5 and 10 DF,  p-value: 0.001193

Analysis of Variance Table

Response: Y
            Df Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)   2 3.5104 1.75521 24.5626 0.0001384
TWI(x1, x2)  1 0.0400 0.04000  0.5598 0.4715843
PQ(x1, x2)   2 0.0325 0.01625  0.2274 0.8006072
Residuals   10 0.7146 0.07146
Lack of fit  3 0.1558 0.05195  0.6508 0.6071993
Pure error   7 0.5588 0.07982
```

On peut voir ici qu'il y a un effet significatif des effet linéaires sur l'appréciation de la note (p.value de FO(x1,x2)=0.006<5%).

Les deux effets (%cacao et année de récolte) influent significativement l'appréciation du chocolat (p.value de x1 et x2 < 5%).

L'effet du cacao a néanmoins une influence plus forte sur l'appréciation du chocolat que l'année de récolte : (|coef(%cacao)|>|coef(annéerécolte)| → 0.61>0.24).

Le coefficient positif et fort du facteur %cacao correspond à un effet positif significatif du facteur sur l'appréciation du chocolat. Plus le chocolat aura un pourcentage de cacao élevé, meilleure il sera.

A l'inverse, le coefficient négatif du facteur « année de récolte » signifie un effet significatif négatif sur l'effet du chocolat. Les chocolats vieux auront tendance à être plus appréciés. L'effet est cependant moins fort que celui du % de cacao.

Les effets quadratiques et l'interaction entre l'année de récolte et le pourcentage de cacao ne sont eux pas significatifs.

L'ajustement global du modèle est bon (lack of fit =0.61). On peut donc rejeter l'hypothèse comme quoi il y a un effet de l'erreur d'ajustement du modèle. Le R^2 est plutôt bon (0.75) ce qui nous confirme la qualité du modèle. On peut s'y fier.

10. **Donnez un couple de paramètres qui optimise Y.**

Ces deux valeurs optimisent Y :

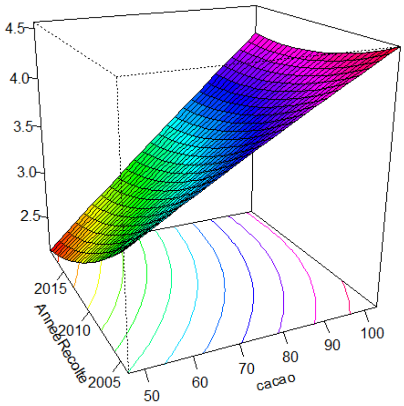
```
Stationary point in original units:
       cacao AnneeRecolte
    229.6159    1989.9052
```

11. **Préciser s'il s'agit d'un point selle, un maximum ou un minimum.**

Les deux valeur propres sont respectivement positives et négatives. Il s'agit donc d'un point selle.

```
Eigenanalysis:
eigen() decomposition
$values
[1]  0.0875 -0.0375
```

12. **Concluez sur l'optimisation de l'appréciation du chocolat selon le % de cacao et l'année de récolte en donnant les combinaisons permettant de maximiser Y pour conseiller l'industriel.**



Le graphique suivant nous permet de modéliser l'appréciation du chocolat selon notre deux variables. On constate que les années de récoltes influencent l'appréciation du chocolat, avec une bonification dans le temps. Plus la fève sera vieille, meilleure sera le chocolat.

Le pourcentage de cacao a lui aussi a un effet sur l'appréciation du chocolat. En effet, plus le pourcentage de cacao est fort, plus le chocolat est apprécié.

Pour conclure, nous conseillerons donc à l'industriel d'opter pour des fèves de cacao plutôt vielles et de maximiser le pourcentage de cacao dans ses tablettes.

Partie 3 : Ouverture et piste de suite

L'industriel a écouté vos conseils. Il va commander des feves plutôt vieilles et opter pour un pourcentage de cacao plutôt fort. Cependant au moment de commander, le fournisseur lui demande de choisir entre des feves venant d'Afrique ou venant d'Amerique du Sud. Il refait appel a vous.

13. **Que lui conseillez vous ? Quelle est votre stratégie ?**

Il faut refaire un plan d'expérience pour déterminer quelle origine de fève optimisera la note d'appréciation du chocolat. A présent, nous sommes face à un plan mixte. Nous avons toujours les deux variables quantitatives (année et %cacao) et nous rajoutons une variable qualitative à deux modalités.

Je choisis la stratégie de me servir de mon plan réaliser pour les deux variables quantitatives en partie 1 et 2 et de le répéter pour chaque modalité des variables qualitatives.

14. **Combien d'essais sont nécessaires ? Présentez la nouvelle matrice des essais.**

2 essais suffisent pour tester les deux modalités d'origine. Nous aurons donc besoin de 2*16 essais = 32 essais. La nouvelle matrice des essais est donnée ci-dessous. Nous réaliserons la même démarche que celle réalisée en partie 1 et 2 pour tester la qualité de ce nouveau plan d'expérience, la modélisation de Y et l'identification de la combinaison des modalités maximisant Y.

```
plan2 <- rbind(plan1,plan1)
plan3 <- cbind(plan2,Origine=c(rep("AmdS",16),rep("Afrique",16)))
```

Exercice 2. Conception d'une nouvelle bière blanche cidrée à l'eau de mer (exemple réel, par Léa Carrolaggi)

L'énoncé de cet exercice repose sur la conception réelle d'une toute nouvelle bière en vue de participer au concours national Def'IAB sur le thème de « élaborer une bière audacieuse ». Après avoir mené une grande réflexion autour du terme « audacieux », les étudiants de l'IUT de Montpellier se sont entendus sur la conception d'une bière blanche cidrée à l'eau de mer. Seulement, le caractère audacieux n'est pas le seul critère évalué lors de ce concours, il y a aussi l'appréciation de la bière. Tout l'enjeu est donc de trouver les bonnes quantités d'eau de mer et de pomme à incorporer dans la bière de façon à ce qu'elle soit agréable en bouche et présente une forte typicité.

Les étudiants décident donc d'organiser une analyse sensorielle sur des échantillons de 10mL de bière blanche industrielle additionnée de différentes quantités de sel (notée X_1) et de différents volumes de jus de pommes (notée X_2) dont les plages de variation s'étendent respectivement de 0,02 à 0,04g et de 2 à 4mL. Un ensemble de juges appréciera chaque échantillon et leur attribuera une note allant de 1 à 10. La moyenne de celles-ci constituera l'appréciation finale de chaque échantillon. L'objectif est de trouver la combinaison {quantité de sel, quantité de jus de pomme} qui conduit à une appréciation maximale de la bière blanche en un minimum de 14 échantillons étant donné que le concours approche à grands pas.

1. **Quel est le nom du plan d'expérience à construire ?**

On s'intéresse ici à l'effet de deux variables explicatives quantitatives à deux modalités sur une variable à expliquer quantitative. Le plan à construire est donc un plan composite centré.

2. **Construire le plan sous R en fixant le paramètre randomize de la fonction que vous utiliserez à FALSE et le paramètre n0 à 3 et en précisant le nom des facteurs ainsi que leur valeur au centre du domaine et leur étendu de variation. Une fois le plan construit, indiquer les expériences correspondant aux points du plan factoriel, aux points en étoile et aux points au centre.**

```
library(rsm) #Chargement du package rsm permettant de construire des plans composites centrés
plancompositecentre <- ccd(2,n0=3,coding=list(x1~(sel-0.03)/0.01,x2~(jus-3)/1),randomize=FALSE)
#Construction du plan composite centré à 2 facteurs avec 6 points au centre (n0*2),les noms des facteurs ainsi que leur valeur au centre du domaine et leur étendu de variation sont précisés dans coding, randomize=FALSE permet de figer la configuration du plan d'expérience
plancompositecentre #Affichage du plan composite centré
```

```
   run.order std.order        sel      jus Block
1          1         1 0.02000000 2.000000     1
2          2         2 0.04000000 2.000000     1   Points du plan factoriel
3          3         3 0.02000000 4.000000     1   complet 2^2
4          4         4 0.04000000 4.000000     1
5          5         5 0.03000000 3.000000     1
6          6         6 0.03000000 3.000000     1   Points au centre
7          7         7 0.03000000 3.000000     1
8          1         1 0.01585786 3.000000     2
9          2         2 0.04414214 3.000000     2   Points en étoile
10         3         3 0.03000000 1.585786     2
11         4         4 0.03000000 4.414214     2
12         5         5 0.03000000 3.000000     2
13         6         6 0.03000000 3.000000     2   Points au centre
14         7         7 0.03000000 3.000000     2
```

3. **Quel est le modèle qui permettra d'analyser les résultats à l'issue des expériences ?**

Le modèle qui est utilisé la plupart du temps pour analyser les résultats d'expériences issues de plans composites centrés est le modèle prenant en compte à la fois les effets linéaires, les effets quadratiques et les effets d'interaction. Il s'écrit de la façon suivante :

$$\forall(i=1,\ldots,n)\ \ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_{11} X_{i1}^2 + \beta_{22} X_{i2}^2 + \beta_{12} X_{i1} X_{i2} + \varepsilon_i$$

Avec $\beta_1 X_{i1}$ et $\beta_2 X_{i2}$ représentant les effets linéaires (effets de X_1 et X_2 seuls), $\beta_{11} X_{i1}^2$ et $\beta_{22} X_{i2}^2$ représentant les effets quadratiques (effets de X_1^2 et X_2^2 seuls), $\beta_{12} X_{i1} X_{i2}$ représentant les effets d'interaction (effet entre X_1 et X_2) et ε_i étant les résidus sur lesquels on émet les hypothèses suivantes :

$$\forall\ (i)\ L(\varepsilon_i) = N\ (0, \sigma)$$
$$\forall\ (i) \neq (i')$$
$$\text{cov}(\varepsilon_i, \varepsilon_i) = 0$$

4. **Construire la matrice X associée à ce plan et au modèle comprenant les effets linéaires, les effets quadratiques et l'effet d'interaction puis calculer la matrice $(X'X)^{-1}$. Interpréter.**

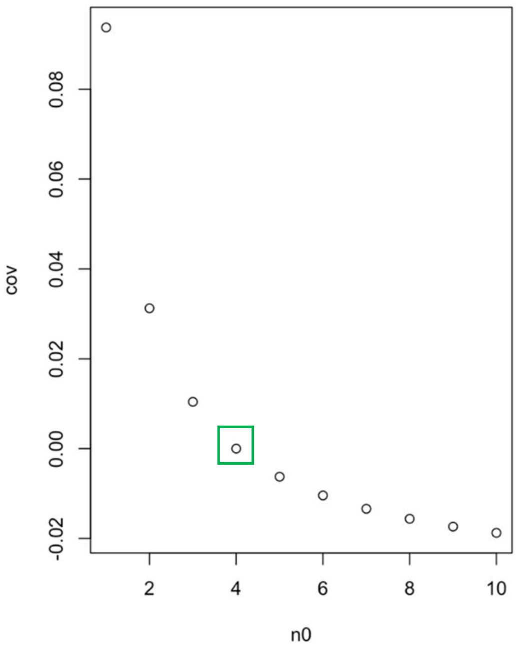
```
X <- model.matrix(~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data=plancompositecentre)
#Création de la matrice des effets comprenant les effets linéaires, les effets quadratiques et l'effet d'interaction
solve(t(X)%*%X) #Affichage de la matrice (X'X)-1
```

```
             (Intercept)    x1    x2     I(x1^2)     I(x2^2) I(x1 * x2)
(Intercept)   0.16666667 0.000 0.000 -0.08333333 -0.08333333       0.00
x1            0.00000000 0.125 0.000  0.00000000  0.00000000       0.00
x2            0.00000000 0.000 0.125  0.00000000  0.00000000       0.00
I(x1^2)      -0.08333333 0.000 0.000  0.13541667  0.01041667       0.00
I(x2^2)      -0.08333333 0.000 0.000  0.01041667  0.13541667       0.00
I(x1 * x2)    0.00000000 0.000 0.000  0.00000000  0.00000000       0.25
```

La constante est parfaitement indépendante des effets principaux et de l'effet d'interaction et est seulement légèrement liée aux effets quadratiques. Les effets principaux et l'effet d'interaction sont indépendants de la constante et de tous les autres effets. Les effets quadratiques sont légèrement liés entre eux et en moindre mesure à la constante. On se retrouve donc avec une matrice $(X'X)^{-1}$ presque diagonale et donc presque parfaite. Tout laisse supposer que les expériences ont été bien choisies et qu'elles vont permettre d'estimer tous les coefficients sans qu'il n'y ait de corrélation entre eux.

5. **Représenter l'évolution de la covariance entre les facteurs quadratiques X_1^2 et X_2^2 en fonction du nombre de points au centre sur R. Combien aurait-il fallu de points au centre pour obtenir une covariance entre ces facteurs nulle ?**

```
n0=1:10 #Création d'un vecteur contenant les valeurs de n0 allant de 1 à 10
cov <- NULL #Création d'une variable à NULL au départ
for (i in 1:length(n0)){ #Pour chaque i allant de 1 à 10 faire la boucle suivante
  plancompositecentre_test <- ccd(2,n0=n0[i],randomize=FALSE,coding=list(x1~(sel-0.03)/0.01,x2~(jus-3)/1)) #Création du plan composite centré en fonction de la valeur de i
  X <- model.matrix(~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data=plancompositecentre_test) #Création de la matrice des effets associé au plan composite centré
  cov <- c(cov,solve(t(X)%*%X)[4,5]) #Crée un vecteur qui récupère toutes les valeurs de covariance se trouvant en ligne 4 colonne 5 de la matrice inversée correspondant à la covariance entre les facteurs quadratiques
}
plot(cov~n0) #Crée un graphe avec en abscisse le nb de pts au centre et en ordonné la covariance
```



D'après le graphique, la covariance est nulle pour n0=4 donc pour 8 points au centre.

6. **Les expériences ont été réalisées et les appréciations globales de chaque échantillon ont été collectées dans un vecteur Y=c(7,4.3,5.5,5,7.6,7.5,7.6,5.3,3,8.2,5.6,7.4,7.5,7.5). Analyser les résultats sous R.**

```
Y <- c(7,4.3,5.5,5,7.6,7.5,7.6,5.3,3,8.2,5.6,7.4,7.5,7.5) #Création d'un vecteur Y contenant les appréciations globales de chaque échantillon
CR.rsm <- rsm(Y~SO(x1,x2),data=plancompositecentre) #Création du modèle comprenant les effets linéaires, les effets quadratiques et l'effet d'interaction grâce au terme SO
summary(CR.rsm) #Affichage des résultats
Call:
rsm(formula = Y ~ SO(x1, x2), data = plancompositecentre)

            Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  7.51667    0.14962  50.2383 2.729e-11 ***
x1          -0.80659    0.12957  -6.2249 0.0002525 ***
x2          -0.55962    0.12957  -4.3189 0.0025499 **
x1:x2        0.55000    0.18325   3.0014 0.0170348 *
x1^2        -1.70208    0.13487 -12.6206 1.459e-06 ***
x2^2        -0.32708    0.13487  -2.4252 0.0415088 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.9661,     Adjusted R-squared:  0.9449
F-statistic: 45.56 on 5 and 8 DF,  p-value: 1.147e-05

Analysis of Variance Table

Response: Y
            Df  Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)   2  7.7100  3.8550 28.7008 0.0002239
TWI(x1, x2)  1  1.2100  1.2100  9.0085 0.0170348
```

```
PQ(x1, x2)    2 21.6797 10.8399 80.7032 4.973e-06
Residuals     8  1.0745  0.1343
Lack of fit   3  1.0462  0.3487 61.5415 0.0002278
Pure error    5  0.0283  0.0057

Stationary point of response surface:
       x1        x2
-0.434128 -1.220468

Stationary point in original units:
       sel        jus
0.02565872 1.77953150

Eigenanalysis:
eigen() decomposition
$values
[1] -0.2741232 -1.7550435

$vectors
         [,1]       [,2]
x1 -0.1891075 -0.9819564
x2 -0.9819564  0.1891075
```

La table de Fisher est la première table à regarder car elle délivre les résultats globaux des effets linéaires, quadratiques et d'interaction. L'effet global des effets linéaires (=FO) est significativement différent de 0 puisque la probabilité critique est inférieure à 5% (égale à 2.24×10^{-4}) ce qui veut dire qu'il y a au moins un des deux facteurs X_1 ou X_2 qui a un effet linéaire significatif sur Y. L'effet d'interaction (=TWI) est significativement différent de 0 puisque la probabilité critique est inférieure à 5% (égale à 1.70×10^{-2}) ce qui veut dire qu'il y a un effet significatif de l'interaction X_1X_2. L'effet global des effets quadratiques (=PQ) est significativement différent de 0 puisque la probabilité critique est inférieure à 5% (égale à 4.97×10^{-6}) ce qui veut dire qu'il y a au moins un des deux facteurs X_1^2 ou X_2^2 qui a un effet linéaire significatif sur Y.

L'erreur d'ajustement mesure l'écart entre les valeurs observées et les valeurs prédites correspondantes. L'erreur pure correspond à la variance de la réponse Y pour les 6 points au centre. L'erreur d'ajustement est significative puisque la probabilité critique est inférieure à 5% (égale à 2.27×10^{-4}) ce qui veut dire qu'elle n'est pas du même ordre de grandeur que l'erreur pure. On ne peut donc pas accepter l'hypothèse H_0 : « il n'y a pas d'erreur d'ajustement » mais il ne semble pas que le modèle soit mal choisi pour autant. L'erreur pure est ici de 0.0057 ce qui est très faible et qui veut dire que les 6 mêmes échantillons ont été perçus de la même manière à chaque fois. Étant donné que le test de Fisher compare l'erreur d'ajustement par rapport à l'erreur pure, le fait que l'erreur pure soit très faible a grandement contribué à rejeter l'hypothèse H_0.

Il y a des effets linéaires, quadratiques et d'interaction significatifs donc il va maintenant être intéressant de regarder les résultats des effets linéaires et quadratiques un par un dans la table de Student. Tous les effets linéaires et tous les effets quadratiques sont significativement différents de 0 puisque leur probabilité critique est à chaque fois inférieure à 5% ce qui veut dire qu'ils ont chacun des effets significatifs sur Y : X_1 et X_2 ont donc des effets linéaires significatifs sur Y, X_1^2 et X_2^2 ont des effets quadratiques significatifs sur Y et X_1X_2 a un effet d'interaction significatif sur Y.

7. Les juges sont-ils bien entraînés ?

```
points_centre = cbind(Y[5],Y[6],Y[7],Y[12],Y[13],Y[14]) #Création d'un vecteur récupérant les valeurs de Y pour les expériences correspondant aux points au centre
moyenne_ecarttype = c(mean(points_centre),sd(points_centre)) #Création d'un vecteur contenant la moyenne ainsi que l'écart-type associé des appréciations globales pour les 6 échantillons constituant les points au centre
moyenne_ecarttype #Affichage du vecteur
[1] 7.51666667 0.07527727
```

Les valeurs de Y pour les expériences n°5, 6, 7, 12, 13 et 14 correspondant aux points au centre ont été récupérées dans un nouveau vecteur à l'aide de la fonction cbind. Il a ensuite été possible de calculer la moyenne des valeurs contenues dans ce vecteur ainsi que l'écart-type associé. On obtient une moyenne de 7.52 et un écart-type de 0.08 ce qui veut dire que lors de la dégustation de ces 6 mêmes échantillons, les juges ont donné une appréciation globale pour ces échantillons de 7.52 avec une très faible variabilité autour de cette moyenne. Ils paraissent donc particulièrement bien entraînés ce qui est rassurant.

8. Trouver la combinaison {quantité de sel, quantité de jus de pomme} qui conduit à une appréciation maximale de la bière blanche sur R. L'enjeu était de taille, les étudiants vous demandent de vérifier manuellement ces valeurs.

Sur R, on trouve l'optimum sans unités dans « stationary point of response surface » : $\{X_1 = -0.4341, X_2 = -1.2205\}$ et l'optimum dans les unités des facteurs dans « stationary point in original units » : $\{X_1 = 0.0257$ g, $X_2 = 1.7995$ mL$\}$.

R nous donne aussi la valeur des coefficients en colonne « Estimate » ce qui nous permet d'écrire le modèle suivant :

$$Y = 7.51667 - 0.80659X_1 - 0.55962 \times X_2 -1.70208 \times X_1^2 - 0.32708 \times X_2^2 + 0.55 \times X_1 \times X_2$$

On peut vérifier l'optimum donné par R manuellement en dérivant l'équation par X_1 puis par X_2 et en annulant les dérivées.

$$\frac{dY}{dX_1}: -\ 0.80659 - 3.40416\, X_1 + 0.55\, X_2 = 0$$
$$\frac{dY}{dX_2}: -\ 0.55962 + 0.65416\, X_2 + 0.55\, X_1 = 0$$

On a ici deux équations à deux inconnues donc on va exprimer X_2 en fonction de X_1 puis intégrer ce terme dans la deuxième équation pour trouver X_1. On en déduira ensuite X_2 en remplaçant X^1 par sa valeur dans la formule de X_2.

$$\frac{\mathrm{dY}}{\mathrm{dX_1}}: -\ 0.80659 - 3.40416\ \mathrm{X_1} + 0.55\ \mathrm{X_2} = 0$$
$$\mathrm{X_2} = \frac{0.80659 + 3.40416\ \mathrm{X_1}}{0.55}$$
$$\frac{\mathrm{dY}}{\mathrm{dX_2}}: -\ 0.55962 + 0.65416 \times \frac{0.80659 + 3.40416\ \mathrm{X_1}}{0.55} + 0.55\ \mathrm{X_1} = 0$$
$$-\ 0.55962 + \frac{0.65416 \times 0.80659}{0.55} + \frac{0.65416 \times 3.40416\ \mathrm{X_1}}{0.55} = -\ 0.55\ \mathrm{X_1}$$
$$-0.55962 + \frac{0.65416 \times 0.80659}{0.55} = -\ 0.55\ \mathrm{X_1} - \frac{0.65416 \times 3.40416\ \mathrm{X_1}}{0.55}$$
$$-0.55962 + \frac{0.65416 \times 0.80659}{0.55} = \mathrm{X_1} \times \left(-0.55 - \frac{0.65416 \times 3.40416}{0.55} \right)$$
$$\mathrm{X_1} = \frac{0.55962 + \frac{0.65416 \times 0.80659}{0.55}}{0.55 - \frac{0.65416 \times 3.40416}{0.55}} = -0.4341$$
$$\mathrm{X_2} = \frac{0.80659 + 3.40416 \times (\text{-}0.4341)}{0.55} = -1.2202$$

On retrouve bien le même optimum sans unités que celui délivré par R à savoir {X_1 = -0.4341, X_2 = -1.2205}. Pour retrouver l'optimum dans les unités des facteurs il suffit de faire un produit en croix à partir des deux relations suivantes :

$$\mathrm{X_1} = \frac{\text{Quantité de sel optimale} - \text{Centre du domaine}}{\text{Étendue}}$$
$$\leftrightarrow \text{Quantité de sel optimale} = \mathrm{X_1} \times \text{ Étendue} + \text{Centre du domaine}$$
$$= -0.4341\ \times\ 0.01 + 0.03$$
$$= 0.0257\ \mathrm{g}$$

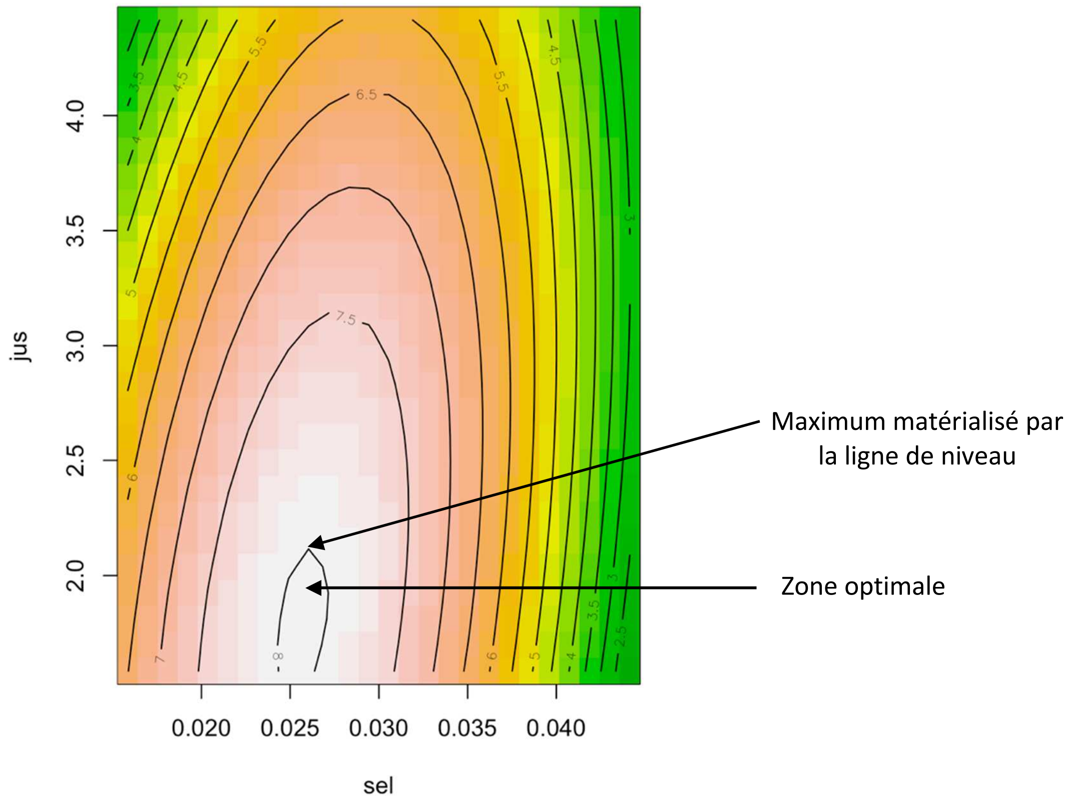
$$\mathrm{X_2} \sim \frac{\text{Volume de jus de pomme optimal} - \text{Centre du domaine}}{\text{Étendue}}$$
$$\leftrightarrow\ \text{Volume de jus de pomme optimal} = \mathrm{X_2} \times \text{ Étendue} + \text{Centre du domaine}$$
$$= -1.2202\ \times\ 1 + 3$$
$$= 1.7798\ \mathrm{mL}$$

On retrouve bien à peu près le même optimum dans les unités des facteurs que celui délivré par R à savoir {X_1 = 0.0257, X_2 = 1.7798}.

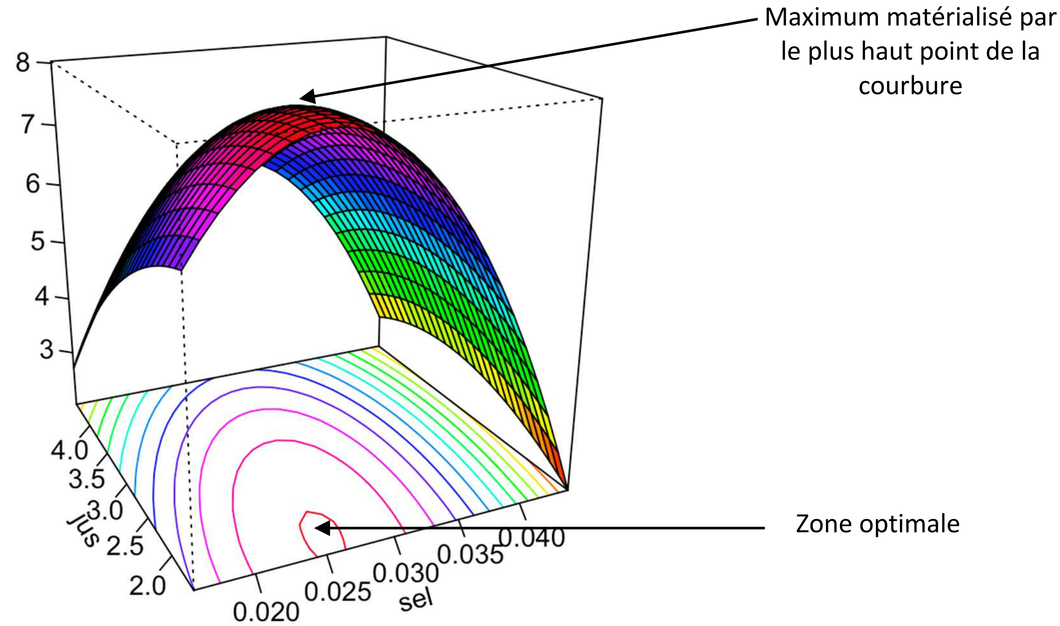
Les valeurs propres délivrées par R dans « Eigenanalysis : eigen() decomposition $values » sont toutes deux négatives et égales à -0.2741 pour X1 et -1.7550 pour X2 ce qui veut dire que l'on a bien affaire à un maximum. Pour que l'échantillon soit agréable en bouche et présente une forte typicité, la quantité de sel optimale est de 0.0257 g et le volume de jus de pomme de 1.7995 mL.

9. **Représenter la surface de réponse Y sur un graphique comprenant des lignes de niveaux puis sur un graphique en 3 dimensions. Indiquer sur chacun d'entre eux le maximum et la zone dans laquelle se trouve les valeurs de X_1 et X_2 optimales.**

```
contour(CR.rsm,~x1+x2,image=TRUE) #Appel de la fonction contour pour tracer le graphique à
lignes de niveaux représentant les résultats de la modélisation contenue dans CR.rsm, ~x1+x2
permet de dire que l'on veut x1 en abscisse et x2 en ordonnée, image=TRUE pour avoir des
couleurs sur le graphique
```




```
persp(CR.rsm,~x1+x2,contours="colors") #Appel de la fonction persp pour tracer le graphique en 3 dimensions représentant les résultats de la modélisation contenue dans CR.rsm, ~x1+x2 permet de dire que l'on veut x1 en abscisse et x2 en ordonnée, contours="colors" pour avoir des couleurs sur le graphique
```



Exercice 3. Formulation d'une crème hydratante ferme et fondante (exemple réel, par Claire Morice)

Une entreprise souhaite simplifier la composition d'une crème (seulement 4 ingrédients : eau, gomme, cire végétale et huile de jojoba) tout en proposant une crème avec une texture **ferme et fondante.**

Remarque : pour des contraintes de formulation, la quantité de cire végétale et d'huile de jojoba ne peuvent pas être modifiées (valeurs fixes).

Quelle quantité d'ingrédients (eau et gomme) et quelle durée de mélange appliquer pour obtenir cette texture souhaitée ?

La variable réponse Y est la mesure de la texture de la crème par un banc de Bostwick (distance en centimètres parcourue par la crème sur une surface en pente, pendant 30 secondes)

Remarque : plus la crème est ferme et fondante, et moins elle s'écoule sur le banc.

Paramètres :

- X1 = ratio eau de raisin/gomme : entre 7,5 et 17,5
- X2 = vitesse du mélange : entre 50 et 150 tours/min

1. **Écrire le modèle utilisé pour analyser ce type de plan généralement**

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon$

Hypothèses sur la résiduelle :

$\forall i = 1,\dots,n$ εi i.i.d. , $E(\varepsilon i)=0$, $V(\varepsilon i)=\sigma^2$ et, $\forall i \neq k$ $cov(\varepsilon_i,\varepsilon_k) = 0$

2. **Construire le plan composite centré sur R : quels sont les essais à réaliser ?**

```
library(rsm)
planB <- ccd(2, n0=2, randomize=FALSE, coding=list (x1~(ratio-12.5)/5, x2~(melange-100)/50))
planB
```

Essai	ratio	melange
1	7.500000	50.00000
2	17.500000	50.00000
3	7.500000	150.00000
4	17.500000	150.00000
5	12.500000	100.00000
6	12.500000	100.00000
7	5.428932	100.00000
8	19.571068	100.00000
9	12.500000	29.28932
10	12.500000	170.71068
11	12.500000	100.00000
12	12.500000	100.00000

```
#en jaune, ce sont les points au centre (= les répétitions)
```

3. **Générer le plan avec les réponses dans R et analyser les résultats avec la fonction rsm. Le modèle est-il bien ajusté ? (R^2 et probabilité critique associée). Quelle est l'erreur pure pour les points au centre ? Quel modèle conservez-vous ? Faire les modifications si nécessaire (effets non significatifs).**

```
library(rsm)
planB <- ccd(2, n0=2, randomize=FALSE, coding=list (x1~(ratio-12.5)/5, x2~(melange-100)/50))
planB
Y<-c(6.3, 3, 9, 5.3, 4, 3.3, 8.3, 2.8, 3.1, 7.5, 4.6, 3.5)
planB.rsm<-rsm(Y~SO(x1,x2),data=planB)
summary(planB.rsm)
```

```
Call:
rsm(formula = Y ~ SO(x1, x2), data = planB)

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  3.85000    0.26800 14.3657 7.123e-06 ***
x1          -1.84727    0.18950 -9.7479 6.698e-05 ***
x2           1.40282    0.18950  7.4026 0.0003122 ***
x1:x2       -0.10000    0.26800 -0.3731 0.7218754
x1^2         0.96875    0.21187  4.5723 0.0038011 **
x2^2         0.84375    0.21187  3.9824 0.0072637 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.9679, Adjusted R-squared:  0.9411
F-statistic: 36.13 on 5 and 6 DF,  p-value: 0.0002101


#seules les interactions ne sont pas significatives

Analysis of Variance Table

Response: Y
            Df Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)   2 43.042 21.5212 74.9099 5.709e-05 #effets linéaires
TWI(x1, x2)  1  0.040  0.0400  0.1392  0.721875 #interaction
PQ(x1, x2)   2  8.823  4.4115 15.3552  0.004366 #effets quadratiques
Residuals    6  1.724  0.2873
Lack of fit  3  0.714  0.2379  0.7067  0.608862 #erreur d'ajustement
Pure error   3  1.010  0.3367                   #erreur pure
```

```
Stationary point of response surface:
        x1         x2 
 0.9133183 -0.7771767 

Stationary point in original units:
   ratio  melange 
17.06659 61.14117 

Eigenanalysis:
eigen() decomposition
$values
[1] 0.9862891 0.8262109             #optimum = un minimum

$vectors
         [,1]       [,2]
x1 -0.9436283 -0.3310069
x2  0.3310069 -0.9436283
```

Le R^2 est très bon (0.968), cela veut dire que l'ajustement global du modèle est excellent. Ce R^2 est très significatif (p value = 0.0002101). Pour les 4 points au centre, la variance vaut 0,3367, cela correspond à l'erreur pure.

Par comparaison des 2 carrées moyens, de l'erreur d'ajustement et de l'erreur pure, on voit que la probabilité critique (0.609) est supérieure à 5%. Cela indique l'erreur d'ajustement est non significative. Le modèle n'est donc pas remis en question.

Concernant le modèle, tous les coefficients sont significatifs, sauf ceux de l'interaction entre x1 et x2. Le modèle est donc conservé sans l'interaction.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} {X_1}^2 + \beta_{22} {X_2}^2 + \varepsilon$$

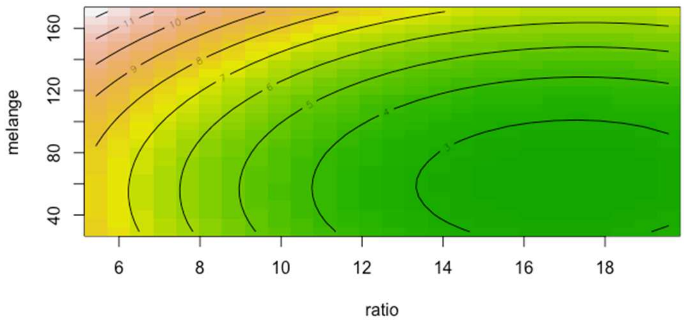
```
library(rsm)
planB <- ccd(2, n0=2, randomize=FALSE, coding=list (x1~(ratio-12.5)/5, x2~(melange-100)/50))
planB
Y <- c(6.3, 3, 9, 5.3, 4, 3.3, 8.3, 2.8, 3.1, 7.5, 4.6, 3.5)
#ici, on note dans le modèle, que les effets linéaire (FO) et les effets quadratiques (PQ)
planB.rsm<-rsm(Y~FO(x1,x2)+PQ(x1,x2),data=planB)
summary(planB.rsm)

Call:
rsm(formula = Y ~ FO(x1, x2) + PQ(x1, x2), data = planB)

            Estimate Std. Error  t value  Pr(>|t|)    
(Intercept)  3.85000    0.25098  15.3398 1.206e-06 ***
x1          -1.84727    0.17747 -10.4089 1.642e-05 ***
x2           1.40282    0.17747   7.9045 9.841e-05 ***
x1^2         0.96875    0.19842   4.8824  0.001789 ** 
x2^2         0.84375    0.19842   4.2524  0.003782 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.9671,	Adjusted R-squared:  0.9483 
F-statistic: 51.46 on 4 and 7 DF,  p-value: 2.829e-05

Analysis of Variance Table

Response: Y
            Df Sum Sq Mean Sq F value   Pr(>F)
FO(x1, x2)   2 43.042 21.5212 85.4129 1.21e-05
PQ(x1, x2)   2  8.823  4.4115 17.5081 0.001887
Residuals    7  1.764  0.2520                 
Lack of fit  4  0.754  0.1884  0.5597 0.711114
Pure error   3  1.010  0.3367                 

Stationary point of response surface:
        x1         x2 
 0.9534306 -0.8312992 

Stationary point in original units:
   ratio  melange 
17.26715 58.43504 

#coordonnées du minimum

Eigenanalysis:
eigen() decomposition
$values
[1] 0.96875 0.84375

#valeurs propres positives : c'est un minimum.
```

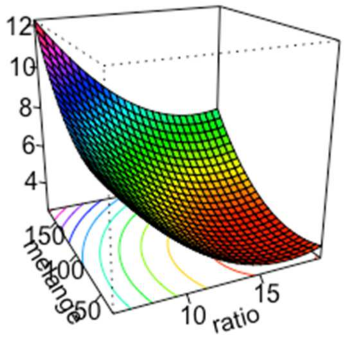
4. **Donner les valeurs des paramètres pour lesquels la texture est optimale (minimum). Donner une estimation de ce minimum.**

La texture est optimale pour un ratio eau de raisin/gomme de 17, et une vitesse de mélange de 58 tours/seconde. Ce point est un minimum car les valeurs propres sont positives. On peut donc mettre environ 26 ml d'eau de raisin et 1,5 g de gomme pour obtenir un ratio de 17.

5. **A l'aide des fonctions contour et persp, représenter les surfaces de réponse.**

```
CR.rsm <- rsm(Y~ FO(x1,x2)+PQ(x1,x2),data=planB)
contour(CR.rsm,~x1+x2,image=TRUE)
persp(CR.rsm,~x1+x2,col=rainbow(50), contours="colors")
```





Exercice 4. Température optimale à cœur des pavés végétaux (exemple fictif, par Anaëlle Lecorgne)

Une entreprise agroalimentaire de pavés végétaux cherche à atteindre une température de cuisson optimale à cœur du produit pour garantir sa sécurité sanitaire : elle doit être supérieure à 82°C. Pour cela, on peut jouer sur le temps de transport du pavé dans le four tunnel (entre 2 et 8 min), noté x1, et le taux d'humidité dans le four (entre 40% et 60%), noté x2. La température du four est fixée à 170°C et ne peut être modifiée.

Nous allons d'abord chercher pour quelles valeurs de x1 et x2 la réponse Y (la température à cœur) est maximum.

1. **Expliciter le modèle utilisé pour analyser ce plan**

$Y = \beta_0+\beta_1X_1+\beta_2X_2+\beta_{11}X_1^2+\beta_{22}X_2^2+\beta_{12}X_1X_2+\xi$

Il s'agit d'un modèle d'analyse de variance à 2 facteurs.

2. **Combien d'essais sont nécessaires à la construction de ce plan ? On prendra 2 points au centre.**

Nombre de points du plan factoriel : n_f = 2k-p = 22= 4

Nombre de points axiaux : n_α = 2k = 2*2 = 4

Nombre de points au centre : n_0 = 2

Nombre total de points : $n_f + n_\alpha + n_0*2 = 12$

3. **Recopier les lignes de code ci-dessuos permettant de générer le plan et de saisir les réponses dans R. Utilisez ensuite la fonction rsm du package rsm pour analyser les résultats. Que pouvez-vous dire sur l'ajustement du modèle (préciser la méthode de déduction) ? Quel modèle conservez-vous ?**
library(rsm)
plan<-ccd(2,n0=2,randomize=FALSE,coding=list(x1 ~(duree-5)/3,x2~(vapeur-50)/10))
Y<-c(72,83,72,86,82,83,65,88,78,79,82,83)

```
library(rsm)
plan<-ccd(2,n0=2,randomize=FALSE,coding=list(x1 ~(duree-5)/3,x2~(vapeur-50)/10))
Y<-c(72,83,72,86,82,83,65,88,78,79,82,83)
plan
plan.rsm<-rsm(Y~SO(x1,x2),data=plan)
summary(plan.rsm)
run.order std.order     duree    vapeur Block
1          1          1 2.0000000 40.00000     1
2          2          2 8.0000000 40.00000     1
```

```
3           3         3 2.0000000 60.00000     1
4           4         4 8.0000000 60.00000     1
5           5         5 5.0000000 50.00000     1
6           6         6 5.0000000 50.00000     1
7           1         1 0.7573593 50.00000     2
8           2         2 9.2426407 50.00000     2
9           3         3 5.0000000 35.85786     2
10          4         4 5.0000000 64.14214     2
11          5         5 5.0000000 50.00000     2
12          6         6 5.0000000 50.00000     2

Data are stored in coded form using these coding formulas ...
x1 ~ (duree - 5)/3
x2 ~ (vapeur - 50)/10

Call:
rsm(formula = Y ~ SO(x1, x2), data = plan)

            Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 82.50000    0.62985 130.9830 1.335e-11 ***
x1           7.19086    0.44537  16.1457 3.589e-06 ***
x2           0.55178    0.44537   1.2389  0.261645
x1:x2        0.75000    0.62985   1.1908  0.278718
x1^2        -2.81250    0.49794  -5.6482  0.001321 **
x2^2        -1.81250    0.49794  -3.6400  0.010835 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.9805,     Adjusted R-squared:  0.9643
F-statistic: 60.42 on 5 and 6 DF,  p-value: 4.741e-05

Analysis of Variance Table

Response: Y
            Df Sum Sq Mean Sq  F value    Pr(>F)
FO(x1, x2)   2 416.10 208.052 131.1095 1.119e-05
TWI(x1, x2)  1   2.25   2.250   1.4179  0.278718
PQ(x1, x2)   2  61.04  30.521  19.2335  0.002457
Residuals    6   9.52   1.587
Lack of fit  3   8.52   2.840   8.5211  0.055929
Pure error   3   1.00   0.333

Stationary point of response surface:
       x1        x2
1.3355128 0.4285273

Stationary point in original units:
    duree    vapeur
 9.006538 54.285273

Eigenanalysis:
eigen() decomposition
$values
[1] -1.6875 -2.9375

$vectors
         [,1]       [,2]
x1 -0.3162278 -0.9486833
x2 -0.9486833  0.3162278
```

L’ajustement global du model est bon (R^2= 0.9643), et il est très significatif (p-value = 4,741e-05). On peut parler d’erreur pure en présence de répétitions, lorsque plusieurs mesures sont faites dans les mêmes conditions expérimentales. Dans le plan composite centré, il s’agit des points au centre. Ici, les 4 valeurs de réponse au centre sont égales à 82, 83, 82 et 83. Pour ces valeurs, la variance vaut 0,333 (Sum Sq/Df = Mean Sq = 0,333) et correspond à l’erreur pure. Pour voir si le modèle a une bonne qualité d’ajustement, on compare les deux carrés moyens de l’erreur pure et de l’erreur d’ajustement par le test de Fisher. La probabilité critique de 0.055929 est supérieure à 5%, et indique un déficit d’ajustement non significatif, on ne remet pas en question le modèle.

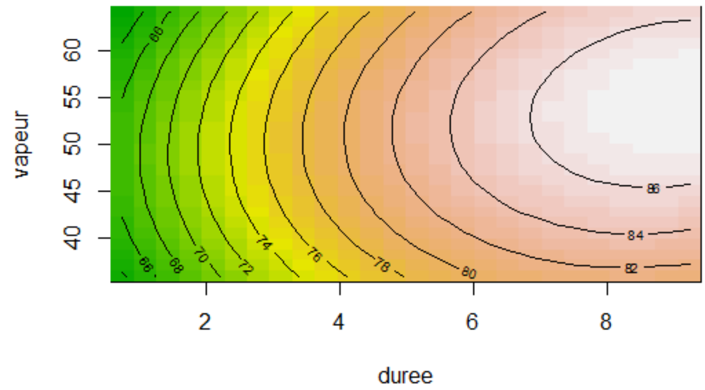
Concernant les variables, on voit dans le tableau d’analyse de variance de la réponse Y, qu’au moins un des coefficients des effets de premier ordre est significatif ainsi qu’au moins un des coefficients des effets quadratiques. Cependant, le coefficient de l’interaction x1x2 n’est pas significatif. Quand on regarde le tableau des variables, on remarque que seul le coefficient de l’effet de 1er ordre x1 est significatif. Concernant les effets quadratiques, les coefficients de x1 et x2 sont significatifs. On pourra ainsi retenir le modèle sans l’interaction et seulement les effets de 1er ordre et les effets quadratiques. Il n’est cependant pas utile de réestimer les coefficients car avec un plan composite centré, l’interaction est orthogonale à tous les autres effets.

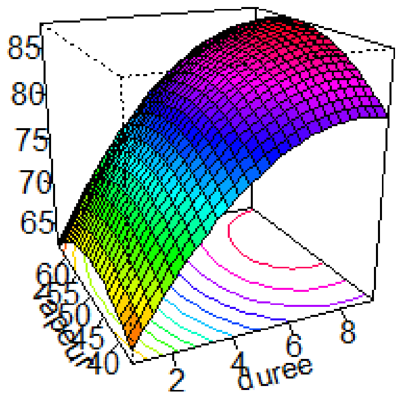
4. **Donner les valeurs de x1, X2 pour lesquels la réponse Y, c'est-à-dire la température à cœur du pavé végétal, est maximale.**

Dans le listage, on peut observer les coordonnées du point stationnaire de la surface de réponse. On voit que les valeurs x1 et de x2 pour lesquelles la température du pavé végétal est à son optimum sont x1= 1,34 et x2=0,43, soit une durée de 9.006538 min et un taux de vapeur de 54.285273 %. Il s'agit bien d'un maximum d'après le « eigenanalysis », dans lequel on voit que les deux valeurs sont négatives.

5. **Afin d'optimiser la cadence de production, on aimerait appliquer un temps de cuisson minimum en atteignant la température minimale de 82°C à cœur du pavé permettant d'assurer sa sécurité sanitaire. Après avoir tracé le graphique des courbes de niveaux de températures en fonction de la durée et de la vapeur, et la surface de réponse de la température d'après le modèle, indiquer le couple (durée, vapeur) avec la durée minimale pour lequel la température de 82°C est atteinte.**

```
contour(plan.rsm,~x1+x2,image=TRUE)
persp(plan.rsm,~x1+x2,col=rainbow(50),contours="colors")
```





Le couple (durée, vapeur) pour lequel la température de 82°C est atteinte avec une durée minimale de passage dans le four semble être approximativement (5 ; 50).

6. **Donner la prévision par votre modèle de la température exacte atteinte pour ce couple (durée, vapeur).**

On prend les valeurs recodées de x1 et x2 :

Y = 82,5 + 7,19086*0 - 2.81250*0^2 -1.81250*0^2 = 82,5

Exercice 5. Caractérisation technologique de céréales extrudées (exemple réel, par Mélanie Demeure)

Lors d'un projet de recherche, on a cherché à développer des céréales extrudées enrichies en inuline, et d'étudier notamment ses caractéristiques physiques, chimique et technologiques. L'extrusion des farines a été faites en variant deux paramètres quantitatifs que sont le pourcentage d'inuline ajoutée et le pourcentage en humidité.

On se propose ici d'étudier l'effet de ces paramètres sur le maintien des céréales dans le lait. Le plan d'expérience a été mis en place avec les lignes de code suivantes :

```
library(rsm)
plan <- ccd(2, n0=2, randomize=FALSE, coding=list (x1~(inuline-15)/7, x2~(humidite-15)/2.5))
```

1. **Expliciter le modèle utilisé pour analyser les résultats d'un tel plan.**

Le modèle complet se compose, pour chacune des variables explicatives d'un effet linéaire, d'un effet quadratique et d'une interaction entre les 2 variables. On a alors la formule suivante :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \epsilon$$

2. **En observant les lignes de codes, que pouvez-vous en déduire sur l'échelle des valeurs testées pour le pourcentage d'inuline et le pourcentage d'humidité ? Combien fait-on de points au centre ?**

Selon la ligne de code, on peut voir que l'on a voulu faire varier le taux d'inuline entre 8 et 22% (15±7), le taux d'humidité entre 12.5 et 17.5% (15±2.5), et inclure 4 points au centre (2 blocs contenant 2 n0 chacun.

3. **Recopiez les lignes de code permettant de générer le plan dans R. On veut finalement remplacer le dernier essai de point au centre par un essai dit « contrôle ». Cela signifie que le pourcentage d'humidité est central, et que le pourcentage d'inuline ajoutée est de 0. Faire les manipulations sur R pour remplacer cet essai.**

Un essai central signifie que ramené à des valeurs centrées, cela donne un essai avec inuline centrée = -15/7 et humidité =0. Les lignes de codes ci-dessous permettent de remplacer ce dernier essai par l'échantillon central.

```
library(rsm)
plan <- ccd(2, n0=2, randomize=FALSE, coding=list (x1~(inuline-15)/7, x2~(humidite-15)/2.5))
plan1 <- plan[1:11,]
plan1 <- rbind(plan1,c(6,6,-15/7,0,2))
```

4. **Saisissez les réponses dans R puis, à l'aide de la fonction `rsm` du package `rsm`, analyser les résultats. Quelle est l'erreur pure ? Que pouvez-vous dire sur l'ajustement du modèle ? Quel modèle conservez-vous ?**
`Y_lait <-c(13.7,5.2,15.3,4.4,4.5,6.0,12.5,5.6,6.9,9.8,5.8,19.7)`

```
Y_lait <- c(13.7,5.2,15.3,4.4,4.5,6.0,12.5,5.6,6.9,9.8,5.8,19.7)
CR.rsm <- rsm(Y_lait~SO(x1,x2),data=plan1)
summary(CR.rsm)

Call:
rsm(formula = Y_lait ~ SO(x1, x2), data = plan1)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  5.74205    0.93493  6.1417 0.0008531 ***
x1          -3.45333    0.57138 -6.0438 0.0009284 ***
x2           0.61265    0.61601  0.9946 0.3583538
x1:x2       -0.60000    0.87116 -0.6887 0.5167233
x1^2         1.58382    0.47846  3.3102 0.0162000 *
x2^2         1.64403    0.73054  2.2504 0.0654036 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.935,      Adjusted R-squared:  0.8808
F-statistic: 17.25 on 5 and 6 DF,  p-value: 0.001675

Analysis of Variance Table

Response: Y_lait
            Df  Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)   2 223.086 111.543 36.7437 0.0004301
TWI(x1, x2)  1   1.440   1.440  0.4744 0.5167233
PQ(x1, x2)   2  37.316  18.658  6.1462 0.0352887
Residuals    6  18.214   3.036
Lack of fit  4  16.888   4.222  6.3647 0.1403684
Pure error   2   1.327   0.663

Stationary point of response surface:
        x1         x2
1.09266564 0.01306153

Stationary point in original units:
 inuline humidite
22.64866 15.03265

Eigenanalysis:
eigen() decomposition
$values
[1] 1.915428 1.312415

$vectors
         [,1]       [,2]
x1 -0.6708780 -0.7415678
x2  0.7415678 -0.6708780
```

Sur ce premier modèle complet, prenant en compte tout les effets, on se rend compte que le modèle a une erreur d'ajustement non significative (p-value de `Lack of fit = 0.1403684).` Le modèle n'est donc pas mal ajusté. Cependant, on se rend compte que les effets liés à l'humidité ne sont pas significatifs. On peut donc refaire un modèle sans interaction, puis sans x2.

```
CR.rsm <- rsm(Y_lait~FO(x1)+PQ(x1),data=plan1)
summary(CR.rsm)

Call:
rsm(formula = Y_lait ~ FO(x1) + PQ(x1), data = plan1)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  7.22938    0.78019  9.2662 6.723e-06 ***
```

```
x1          -3.49750    0.67373 -5.1912 0.0005707 ***
x1^2         1.20338    0.52809  2.2787 0.0486628 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.8642,      Adjusted R-squared:  0.834
F-statistic: 28.64 on 2 and 9 DF,  p-value: 0.0001253

Analysis of Variance Table

Response: Y_lait
            Df  Sum Sq Mean Sq F value    Pr(>F)
FO(x1)       1 220.083 220.083 52.0826 4.992e-05
PQ(x1)       1  21.942  21.942  5.1927   0.04866
Residuals    9  38.031   4.226
Lack of fit  3  20.691   6.897  2.3865   0.16783
Pure error   6  17.340   2.890

Stationary point of response surface:
      x1
1.453201

Stationary point in original units:
inuline
25.1724

Eigenanalysis:
eigen() decomposition
$values
[1] 1.203379

$vectors
   [,1]
x1    1
```

On se retrouve finalement avec un modèle bien ajusté ($R^2 = 0.8642$), et la p-value associée, égale à 0.0001253, nous permet de conclure à la significativité du R^2. L'erreur pure vaut 2.890. La probabilité critique (0.16783) indique un déficit d'ajustement non significatif. On peut donc garder ce dernier modèle.

5. **Ecrire le modèle finalement conservé pour la recherche de l'optimum et donner la valeur du couple (X1, X2) pour lequel la variable Y est optimum. Donner une estimation de cet optimum. Est-ce un maximum ou un minimum ?**

Le modèle conservé est le modèle sans interaction.

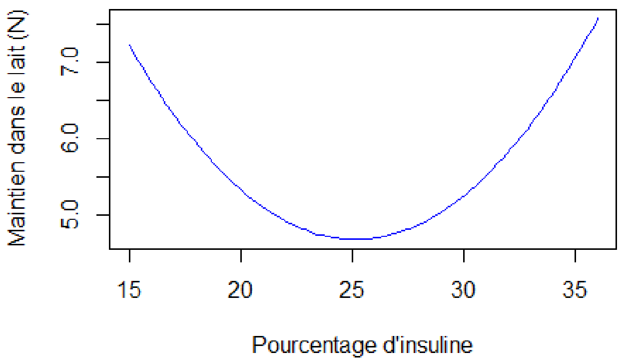
$$Y = 7.22938 - 3.49750x_1 + 1.20338x_1^2$$

L'optimum se trouve au point d'abscisse 1.45 en unités centrées, au point de coordonnées 25.17 en unités originales. Il vaut alors 4.688. La valeur propre (*eingenvalue*) est positive donc c'est un minimum.

6. **Représentez graphiquement les réponses.**

On utilise les lignes de code suivantes.

```
x <- seq(0,3,length.out = 100)
y <-CR.rsm$coefficients[1] + CR.rsm$coefficients[2]*x + CR.rsm$coefficients[3]*x^2
plot(x*7+15,y,col="blue",ylab="Maintien dans le lait (N)",xlab="Pourcentage d'insul d'insuline",
type="l")
```



Exercice 6. Etude de l'irrigation et de la fertilisation (exemple fictif, par E. David)

L'exercice suivant se base sur une thématique et des données fictives. Un chercheur étudie l'effet du stress hydrique et de la carence en azote sur une variable d'intérêt : la production de maïs. Il doit ainsi étudier l'effet de deux facteurs sur le rendement en maïs grain en quintaux par hectare. Le chercheur ne peut réaliser que 12 essais. Il cherche ainsi la combinaison des facteurs conduisant à une valeur maximum de la réponse Y. Les facteurs quantitatifs sont :

- F1 : Quantité d'eau irriguée (2 modalités : 230 mm/ha et 280 mm/ha)
- F2 : Quantité d'azote fertilisée (2 modalités : 100 kg d'N/ha et 200 kg d'N/ha)

Les données choisies ici ne sont pas de vraies données scientifiques issues de réelles expérimentations. Les modalités choisies et les rendements présentés sont néanmoins cohérents avec les valeurs utilisées et observées en agriculture aujourd'hui.

1. **Générer le plan associé à cette expérience sous R.**

On construit un plan composite centré en 12 essais avec 4 essais au centre.

```
# Création du plan d'expérience composite centré
library(rsm)
plan = ccd(2, n0=2, randomize=FALSE, coding=list (x1~(Irrigation-255)/25, x2~(Fertilisation-
150)/50))
plan

run.order std.order Irrigation Fertilisation Block
1          1         1  230.0000     100.00000     1
2          2         2  280.0000     100.00000     1
3          3         3  230.0000     200.00000     1
4          4         4  280.0000     200.00000     1
5          5         5  255.0000     150.00000     1
6          6         6  255.0000     150.00000     1
7          1         1  219.6447     150.00000     2
8          2         2  290.3553     150.00000     2
9          3         3  255.0000      79.28932     2
10         4         4  255.0000     220.71068     2
11         5         5  255.0000     150.00000     2
12         6         6  255.0000     150.00000     2

Data are stored in coded form using these coding formulas ...
x1 ~ (Irrigation - 255)/25
x2 ~ (Fertilisation - 150)/50
```

2. **Analyser les résultats du plan construit à la question 1.**

```
# Analyse des résultats du plan
Y = c(72, 80, 94, 96, 86, 87, 82, 89, 75, 97, 88, 87)
CR.rsm = rsm(Y~SO(x1,x2),data=plan)
summary(CR.rsm)

Call:
rsm(formula = Y ~ SO(x1, x2), data = plan)

            Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 87.00000    0.57935 150.1672 5.882e-12 ***
x1           2.48744    0.40967   6.0719  0.000906 ***
x2           8.63909    0.40967  21.0882 7.410e-07 ***
x1:x2       -1.50000    0.57935  -2.5891  0.041262 *
x1^2        -0.81250    0.45802  -1.7739  0.126432
x2^2        -0.56250    0.45802  -1.2281  0.265391
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.988,      Adjusted R-squared:  0.9779
F-statistic: 98.44 on 5 and 6 DF,  p-value: 1.131e-05

Analysis of Variance Table

Response: Y
            Df Sum Sq Mean Sq  F value    Pr(>F)
FO(x1, x2)   2 646.57  323.28 240.7892 1.863e-06
TWI(x1, x2)  1   9.00    9.00   6.7034   0.04126
PQ(x1, x2)   2   5.29    2.65   1.9707   0.21985
Residuals    6   8.06    1.34
Lack of fit  3   6.06    2.02   3.0278   0.19360
Pure error   3   2.00    0.67

Stationary point of response surface:
       x1        x2
```

```
 24.08359 -24.43226

Stationary point in original units:
   Irrigation Fertilisation
     857.0897    -1071.6132

Eigenanalysis:
eigen() decomposition
$values
[1]  0.07284532 -1.44784532

$vectors
         [,1]       [,2]
x1  0.6463749 -0.7630200
x2 -0.7630200 -0.6463749
```

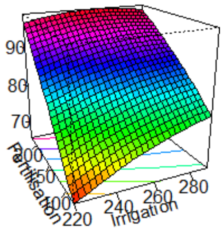
L'ajustement global du modèle est bon avec une valeur de $R^2 = 0{,}988$. La probabilité critique est p-value = 1.131. 10^{-5} donc l'ajustement global du modèle aux données est significatif.
Toutes les mesures sont réalisées dans les mêmes conditions expérimentales : même climat, même type de sol, même variété ... soit des conditions biotiques et abiotiques similaires. On peut donc étudier l'erreur pure du modèle. Elle vaut 0.67 pour ce modèle.
La p-value associée à l'erreur d'ajustement du modèle vaut 0.19360. On accepte donc l'hypothèse HO disant que le modèle est bien ajusté. Il n'y a pas d'erreur d'ajustement donc notre modèle est acceptable.
D'après le premier tableau, il y a un effet significatif des effets linéaires x1 et x2, et un effet significatif de l'interaction x1 : x2 au seuil 5 %. Les effets quadratiques ne sont pas significatifs.
D'après le tableau 2 d'analyse de variance, tous les coefficients sont significatifs à l'exception des coefficients des effets quadratiques $x1^2$ et $x2^2$ au seuil de 5 %. On peut donc retirer les effets quadratiques du modèle et conserver les effets linéaires et d'interaction.

3. **Présenter le modèle conservé après l'analyse de la question 2 ainsi que la surface de réponse. Donner une estimation de l'optimum atteint par le modèle.**

On construit un nouveau modèle en retirant les effets quadratiques car ils ne sont pas significatifs d'après les résultats de l'analyse de variance réalisée à la question 2.

```
# Modèle sans effet quadratique
CR.rsm = rsm(Y~FO(x1,x2)+TWI(x1,x2),data=plan)
summary(CR.rsm)
Call:
rsm(formula = Y ~ FO(x1, x2) + TWI(x1, x2), data = plan)

            Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 86.08333    0.37287 230.8650 < 2.2e-16 ***
x1           2.48744    0.45667   5.4469  0.000611 ***
x2           8.63909    0.45667  18.9174 6.306e-08 ***
x1:x2       -1.50000    0.64584  -2.3226  0.048723 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:   0.98,      Adjusted R-squared:  0.9726
F-statistic:   131 on 3 and 8 DF,  p-value: 3.87e-07

Analysis of Variance Table

Response: Y
            Df Sum Sq Mean Sq  F value    Pr(>F)
FO(x1, x2)   2 646.57  323.28 193.7679 1.673e-07
TWI(x1, x2)  1   9.00    9.00   5.3944   0.04872
Residuals    8  13.35    1.67
Lack of fit  5  11.35    2.27   3.4042   0.17111
Pure error   3   2.00    0.67

Stationary point of response surface:
      x1       x2
5.759392 1.658291

Stationary point in original units:
   Irrigation Fertilisation
     398.9848      232.9146

Eigenanalysis:
eigen() decomposition
$values
[1]  0.75 -0.75
```

```
$vectors
          [,1]       [,2]
x1 -0.7071068 -0.7071068
x2  0.7071068 -0.7071068

# Représentation de la surface de réponse
contour(CR.rsm,~x1+x2,image=TRUE)
persp(CR.rsm,~x1+x2,contours="colors")
```



Après correction, le modèle est : **Y = 86.08 + 2.48 x1 + 8.63 x2 – 1.50 x1.x2**

Pour connaitre le couple optimal (x1, x2) on dérive la variable réponse Y par rapport à x1 et x2 :

- dY/dx1 = 2.48 – 1.50.x2= 0 ➔ x2 = 2.48/1.50 = 1.65
- dY/dx2 = 8.63 – 1.50 .x1= 0 ➔ x1 = 8.63/1.50 = 5.75

Ainsi on a x1 = 5.75 et x2 = 1.65 ce qui correspond en unité originale à x1 = 398.98 mm/ha et x2 = 232.91 kg/ha. D'après le signe des valeurs propres associées, le point stationnaire décrit est un point selle.

Ce modèle ne permet pas de déterminer un couple optimal (x1, x2) réalisant un maximum de rendement, seulement un point selle est mis en évidence. L'hypothèse du chercheur quant à l'existence d'effets quadratiques des deux facteurs étudiés n'est pas vérifiée.

Néanmoins, ces résultats ne se basent pas sur des données réelles de terrain mais sur des données fictives. On ne peut donc pas conclure quant à l'existence d'effets quadratiques de l'irrigation et de la fertilisation azotée en réalité. Cependant, dans cette simulation de jeu de données, nous avons quand même mis en évidence une interaction entre l'irrigation et la fertilisation azotée. L'effet de la variable irrigation dépend donc de la variable fertilisation azotée ce qui est vrai en agronomie. En effet, l'efficacité des plantes à absorber l'eau issue de l'irrigation dépend de sa condition nutritionnel (de son état de santé) qui est directement liée à la quantité d'azote disponible pour la plante par exemple. Ainsi, il y aura un impact sur le rendement final.

Exercice 7. Cuisson optimale d'un œuf (exemple fictif, par Clarisse Thiard)

On cherche à déterminer le temps de cuisson optimal d'un œuf. Pour cela, on considère deux facteurs, le temps de cuisson et la puissance de cuisson (notés X1 et X2). La variable réponse Y est la rigidité de l'œuf obtenu.

1. **Quels sont les différents effets à prendre en compte lors de la planification expérimentale pour 2 facteurs ?**

On doit prendre en compte trois types d'effets :

- les effets linéaires β1 et β2
- les effets quadratiques β11 et β22
- l'effet de l'interaction des deux facteurs β12

Le modèle s'exprime ainsi :

$Y= \beta 0 + \beta 1X1 + \beta 2X2 + \beta 11X1^2 + \beta 22X2^2 +B12X1X2$

2. **Proposer une ligne de code permettant de générer le plan de manière aléatoire suivant les effets donnés en comprenant le temps de cuisson entre 5 et 25 min et la température de 100°C à 200°C.**

```
library(rsm)
plan <- ccd(2, n0=2, randomize=TRUE, coding=list (x1~(time-15)/10, x2~(temperature-150)/50))
```

3. **On considère maintenant un plan non aléatoire et on ajout la commande R : Y<-c(1,2,3,4,5,6,7,8,9,10,11,12). Déterminer les valeurs propres.Définir le type de ces valeurs propres.**

On effectue la commande

```
oeuf.rsm <- rsm(Y~SO(x1,x2),data=plan)
summary(oeuf.rsm)
```

On obtient ainsi deux valeurs propres,

- x1 = -2,463307^e-16 et x2 = -1, les deux valeurs sont négatives, il s'agit donc d'un maximum

- x1=-1 et x2=2,463307^e-16, les deux valeurs ne sont pas du même signe, il s'agit donc d'un point selle.

4. **Que pouvez vous déduire des différents coefficients obtenus ?**

Voici les différents coefficients obtenus :

```
              Estimate  Std. Error t value Pr(>|t|)
(Intercept)  8.5000e+00  2.1397e+00  3.9725 0.007346 **
x1           4.2678e-01  1.5130e+00  0.2821 0.787363
x2           6.7678e-01  1.5130e+00  0.4473 0.670346
x1:x2        4.9266e-16  2.1397e+00  0.0000 1.000000
x1^2        -2.0000e+00  1.6916e+00 -1.1823 0.281801
x2^2        -1.0000e+00  1.6916e+00 -0.5912 0.575987
---
```

Une seule valeur est significative, celle de la moyenne, on peut donc dire que le plan n'est as adapté à l'expérience. On ne peut donner les valeurs des différents β autres que β0.

Exercice 8. La fermentation du vin (exemple fictif, par Baptiste Turban)

La fabrication du vin suit un process de fabrication bien précis. Pour obtenir le vin que désire l'oenologue une étape en particulier doit être maitrisée: la fermentation. Lors de cette étape, de l'acide volatile est créée grâce à l'action des levures. Si cette concentration en acide est trop élevée (supérieur à 0.9g/L) alors le vin n'est pas commercialisable. On voit bien ici tout l'intérêt de rechercher une concentration en acide volatile faible pour limiter le risque de perte. Dans cette étape, deux facteurs (notés X1 et X2) ont été identiés comme importants : la Température et la quantité de levures introduite. La température doit être comprise entre 15 et 25°C. La quantité de levures est comprise entre 20 et 40 g/hl. On recherche pour quelles valeurs de X1 et X2 la réponse Y (concentration en acide volatile) est minimum.

1. **On souhaite au maximum réaliser 12 expériences. Construire un plan d'expérience et donner le nombre de points au centre.**

Dans notre cas nous avons 2 variables quantitatives.En utilisant la formule: Nb expérience = 2^k + 2k + n0 avec k=2, on trouve 4 essais pour le plan complet, 4 essais pour les points en étoile et donc 4 essais pour les points au centre

```
library(rsm)
plan <- ccd(2, n0=2, randomize=FALSE, coding=list (x1~(temp-20)/5, x2~(lev-30)/10))
plan

    run.order std.order     temp      lev Block
 1          1         1 15.00000 20.00000     1
 2          2         2 25.00000 20.00000     1
 3          3         3 15.00000 40.00000     1
 4          4         4 25.00000 40.00000     1
 5          5         5 20.00000 30.00000     1
 6          6         6 20.00000 30.00000     1
 7          1         1 12.92893 30.00000     2
 8          2         2 27.07107 30.00000     2
 9          3         3 20.00000 15.85786     2
 10         4         4 20.00000 44.14214     2
 11         5         5 20.00000 30.00000     2
 12         6         6 20.00000 30.00000     2

 Data are stored in coded form using these coding formulas ...
 x1 ~ (temp - 20)/5
 x2 ~ (lev - 30)/10
```

2. **Décrire le modèle utilisé pour analyser les résultats d'un tel plan.**

Classiquement, pour chaque variable explicative, un effet linéaire et un effet quadratique,auxquels s'ajoute un terme d'interaction entre X1 et X2 :

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_{11} X1^2 + \beta_{22} X2^2 + \beta_{12} X1X2 + \varepsilon$$

3. **Commentez la qualité d'un tel plan:**

```
X = model.matrix(~SO(x1,x2),data=plan)
solve(t(X)%*%X)
```

```
              (Intercept) SO(x1, x2)x1 SO(x1, x2)x2 SO(x1, x2)x1:x2 SO(x1, x2)x1^2 SO(x1, x2)x2^2
(Intercept)         0.250        0.000        0.000            0.00       -0.12500       -0.12500
SO(x1, x2)x1        0.000        0.125        0.000            0.00        0.00000        0.00000
SO(x1, x2)x2        0.000        0.000        0.125            0.00        0.00000        0.00000
SO(x1, x2)x1:x2     0.000        0.000        0.000            0.25        0.00000        0.00000
SO(x1, x2)x1^2     -0.125        0.000        0.000            0.00        0.15625        0.03125
SO(x1, x2)x2^2     -0.125        0.000        0.000            0.00        0.03125        0.15625
```

Ici on a une matrice qui est presque diagonale, la constante dépend très légèrement des effet aux carrés, X1 et X2 sont complétement indépendants, l'effet d'intéraction est indépendant mais les effets au carré eux dépendent de la constante, ils ne sont pas vraiment indépendants. Pour avoir l'orthogonalité complète il faudrait rajouter des expériences.

4. **Au vu des résultats, le patron de la cave décide d'ajouter 4 essais supplémentaires. Les essaies sont ensuite réalisés. Analyser ces résultats. Quelle est l'erreur d'ajustement ? Que pouvez-vous dire sur l'ajustement du modèle ? commentez**

```
plan2= ccd(2, n0=4, randomize=FALSE, coding=list (x1~(temp-20)/5, x2~(lev-30)/10))
Y = c(0.54,0.6,0.42,0.92,0.52,0.53,0.45,0.45,0.2,0.96,0.32,0.59,0.52,0.46,0.53,0.45)
CR.rsm <- rsm(Y~SO(x1,x2),data=plan2)
summary(CR.rsm)
```

```
 Call:
 rsm(formula = Y ~ SO(x1, x2), data = plan2)

             Estimate Std. Error t value  Pr(>|t|)    
(Intercept) 0.488750   0.029360 16.6468 1.280e-08 ***
x1          0.204350   0.029360  6.9602 3.899e-05 ***
x2          0.072730   0.029360  2.4772   0.03270 *  
x1:x2       0.110000   0.041521  2.6492   0.02434 *  
x1^2        0.071250   0.029360  2.4268   0.03565 *  
x2^2        0.008750   0.029360  0.2980   0.77178    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.8711, Adjusted R-squared:  0.8066 
F-statistic: 13.52 on 5 and 10 DF,  p-value: 0.0003521

Analysis of Variance Table

Response: Y
            Df  Sum Sq  Mean Sq F value    Pr(>F)
FO(x1, x2)   2 0.37639 0.188195 27.2901 8.902e-05
TWI(x1, x2)  1 0.04840 0.048400  7.0185  0.024345
PQ(x1, x2)   2 0.04122 0.020612  2.9890  0.096025
Residuals   10 0.06896 0.006896                  
Lack of fit  3 0.05827 0.019424 12.7224  0.003199
Pure error   7 0.01069 0.001527                  

Stationary point of response surface:
        x1         x2 
-0.4605479 -1.2611111 

Stationary point in original units:
    temp      lev 
17.69726 17.38889 

Eigenanalysis:
eigen() decomposition
$values
[1]  0.1032579 -0.0232579

$vectors
         [,1]       [,2]
x1 -0.8642943  0.5029864
x2 -0.5029864 -0.8642943
```

Ici les expériences n'ont pas vraiment été menées, donc les résultats sont incohérents. La probabilité critique de l'erreur d'ajustement(Lack of fit) est significative ce qui veut dire que l'écart entre les points des conditions expérimentales et les points du modèle est significativement plus grand que ce à quoi on peut s'attendre du seul fait de la variabilité résiduelle. Ainsi le modèle est rejeté.

Si les expériences avaient été menées, il aurait fallu regarder si le modèle était bien ajusté avec le Adjusted-R et si tous les effets étaient significatifs. Dans le cas, où un effet n'était pas significatif il aurait été enlevé et un modèle sans cet effet aurait *été construit.Il aurait fallu aussi regarder la p-value de l'erreur d'ajustement pour voir si le modèle était à remettre en question.*

Dans le cas où le modèle était ajusté, on aurait regarder le signe des valeurs propres(Stationary point of response surface). Si les deux signes étaient positifs alors le minimun de concentration en acide volatile aurait été atteint pour les valeurs temp et lev affichées.

Exercice 9. Effet du temps de trempage dans l'acide salicylique sur le pH de tomates (exemple fictif, par Junyi Zhao)

On cherche à déterminer l'effet du temps de trempage de dans l'acide salicylique sur le pH de tomates

Facteurs :

- X1 : Concentration en acide salicylique de 0mM à 4 mM
- X2 : temps de trempage 10 à 30 min

Variable d'intérêt Y : pH de la tomate

1. **Quel est le modèle classique pour analyser ce plan continu ?**

$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon_i$

```
library(rsm)
plan <- ccd(2, randomize = FALSE, coding = list (x1~(Concentration-2)/2, x2~(Temps-20)/10))
plan
Y <- c(1,2,3,4,5,6,7,8,1,2,3,4,5,6,7,8)
```

2. **2. Analyser ce plan avec le package rsm. Commenter l'ajustement du modèle.**

```
library(rsm)
plan <- ccd(2, randomize = FALSE, coding = list (x1~(Concentration-2)/2, x2~(Temps-20)/10))
plan
Y <- c(1,2,3,4,5,6,7,8,1,2,3,4,5,6,7,8)
CR.rsm <- rsm(Y~SO(x1,x2), data = plan)
summary(CR.rsm)

Call :
rsm(formula = Y ~ SO(x1, x2), data = plan)

               Estimate  Std. Error t value  Pr(>|t|)
(Intercept)  6.5000e+00  3.6876e-01 17.6267 7.351e-09 ***
x1           4.2678e-01  3.6876e-01  1.1573  0.274033
x2           6.7678e-01  3.6876e-01  1.8353  0.096340 .
x1:x2       -4.9266e-16  5.2150e-01  0.0000  1.000000
x1^2        -2.5000e+00  3.6876e-01 -6.7795 4.862e-05 ***
x2^2        -1.5000e+00  3.6876e-01 -4.0677  0.002259 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.8705,     Adjusted R-squared:  0.8057
F-statistic: 13.44 on 5 and 10 DF,  p-value: 0.0003601

Analysis of Variance Table

Response: Y
            Df Sum Sq Mean Sq F value   Pr(>F)
FO(x1, x2)   2  5.121   2.561  2.3538   0.1453
TWI(x1, x2)  1  0.000   0.000  0.0000   1.0000
PQ(x1, x2)   2 68.000  34.000 31.2538 4.99e-05
Residuals   10 10.879   1.088
Lack of fit  3  0.879   0.293  0.2050   0.8898
Pure error   7 10.000   1.429

Stationary point of response surface:
        x1         x2
0.08535534 0.22559223

Stationary point in original units:
Concentration         Temps
     2.170711     22.255922

Eigenanalysis:
eigen() decomposition
$values
```

```
[1] -1.5 -2.5

$vectors
            [,1]          [,2]
x1  2.463307e-16 -1.000000e+00
x2 -1.000000e+00 -2.463307e-16
```

```
CR.rsm <- rsm(Y~FO(x1,x2)+PQ(x1,x2), data = plan)

Summary(CR.rsm)
```

```
Call:
rsm(formula = Y ~ FO(x1, x2) + PQ(x1, x2), data = plan)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  6.50000    0.35160 18.4870 1.241e-09 ***
x1           0.42678    0.35160  1.2138  0.250235
x2           0.67678    0.35160  1.9249  0.080486 .
x1^2        -2.50000    0.35160 -7.1104 1.967e-05 ***
x2^2        -1.50000    0.35160 -4.2662  0.001329 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.8705,     Adjusted R-squared:  0.8234
F-statistic: 18.48 on 4 and 11 DF,  p-value: 7.588e-05

Analysis of Variance Table

Response: Y
            Df Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)   2  5.121   2.561  2.5892    0.1198
PQ(x1, x2)   2 68.000  34.000 34.3792 1.853e-05
Residuals   11 10.879   0.989
Lack of fit  4  0.879   0.220  0.1538    0.9552
Pure error   7 10.000   1.429

Stationary point of response surface:
        x1         x2
0.08535534 0.22559223

Stationary point in original units:
Concentration         Temps
     2.170711     22.255922

Eigenanalysis:
eigen() decomposition
$values
[1] -1.5 -2.5

$vectors
   [,1] [,2]
x1    0   -1
x2   -1    0
```

L’ajustement est correct (R^2 = 0.80) et significatif.

Les coefficients de premier ordre et d’interaction ne sont pas significatifs, on enlève du modèle l’effet d’interaction.

3. **Donner l’estimation du modèle avec ses coefficients. Estimer le couple (X1,X2) pour lequel Y est un optimum.**

On a : $Y = 6.5 + 0.43\ X_1 + 0.68\ X_2 - 2.5\ X_1^2 - 1.5\ X_2^2$

On a : X_1 = 0.09 mM et X_2 = 0.23 min

Pour ces valeurs, Y = 6.60

4. **Y est-il un maximum ou un minimum ?**

```
contour(CR.rsm,~x1+x2, image = TRUE)
persp(CR.rsm,~x1+x2, col=rainbow(50), contours = "colors")
```

Il s’agit d’un maximum.

Exercices sur les plans avancés

Exercice 1. Optimisation de la culture de plants de tomates (exemple fictif, par Manon Filippi)

Une jardinerie spécialisée dans la culture biologique cherche à optimiser la culture de ses plants de tomates sous serre pour maximiser la taille des plants 8 semaines après sorti de terre, date à laquelle les particuliers pourront les acheter et les repiquer dans leur jardin. La taille des plants (évaluée en cm) semble dépendre principalement de 4 facteurs :

- F1. la température de la serre : il y a 4 températures différentes (T1, T2, T3, T4) ;
- F2. l'ensoleillement : il y a 3 programmes de luminosité possibles pour la serre (L1,L2,L3) ;
- F3. l'humidité : il y a 2 systèmes d'arrosage (H1,H2) ;
- F4. la quantité d'engrais riche en phospho-potassium apportée : il y a 4 doses d'engrais possibles (D1, D2, D3, D4).

1. **Quel nombre d'essais doit-on réaliser si l'on veut étudier le plan complet ? Construire ce plan complet sous R.**

Le nombre maximum d'essais est égal à $2x3x4^2$ = 96 essais pour le plan complet.

```
library(DoE.base)
plan <- fac.design(nlevels=c(2,3,4,4))
plan
   A B C D
1  2 2 3 3
2  2 2 4 1
3  2 1 1 3
4  2 2 2 3
5  1 1 2 1
6  2 2 3 1
7  1 2 1 2
8  1 2 4 3
9  2 2 4 3
10 2 3 1 1
11 2 2 1 3
12 1 2 3 1
13 2 3 2 1
14 2 1 2 2
15 1 3 2 1
16 1 3 2 4
17 2 2 2 1
18 2 3 4 2
19 1 3 3 3
20 1 3 4 1
21 2 2 1 1
22 2 1 4 1
23 2 3 2 2
24 1 2 3 2
25 1 2 4 2
26 2 1 4 4
27 2 3 3 4
28 1 2 2 1
29 1 3 2 2
30 2 1 4 3
31 1 1 4 1
32 1 3 1 4
33 2 2 4 4
34 1 1 2 4
35 1 2 2 2
36 1 1 4 2
37 1 1 3 1
38 2 1 3 3
39 2 1 3 2
40 1 2 1 1
41 1 1 1 4
42 2 3 4 1
43 2 3 1 2
44 2 1 3 1
45 1 1 2 2
46 2 3 1 4
47 1 3 1 1
48 1 2 1 4
49 1 2 3 4
50 1 1 1 3
51 1 2 4 4
52 2 3 2 3
53 2 2 4 2
54 1 3 4 2
55 1 2 1 3
56 1 1 2 3
57 2 3 3 2
58 1 3 3 4
59 1 2 3 3
60 2 3 3 3
61 1 1 3 3
62 2 2 2 2
63 2 1 4 2
64 2 1 3 4
65 2 2 3 2
66 1 3 1 2
67 1 3 4 3
68 1 1 3 2
69 1 3 1 3
70 2 3 2 4
71 1 3 2 3
72 1 2 2 3
73 1 1 4 3
74 2 1 2 1
75 1 3 4 4
76 1 1 4 4
77 2 1 1 2
78 1 3 3 1
79 2 3 4 4
80 1 1 3 4
81 2 3 1 3
82 2 3 4 3
83 2 1 1 4
84 2 2 1 4
85 2 1 2 3
86 2 3 3 1
87 1 1 1 1
88 2 2 2 4
89 1 2 4 1
90 2 2 1 2
91 2 1 2 4
92 1 3 3 2
93 2 2 3 4
94 2 1 1 1
95 1 2 2 4
96 1 1 1 2
class=design, type= full factorial
```

2. **Combien d'essais faut-il au minimum pour avoir l'orthogonalité entre les facteurs 2 à 2 ? Construire le plan orthogonal sous R.**

Pour l'orthogonalité des facteurs il faut :

F1⊥F2 et F2⊥F4 donc minimum 4x3 = 12 essais

F1⊥F3 et F3⊥F4 donc minimum 4x2= 8 essais

F1⊥F4 donc minimum 4x4 = 16 essais

F2⊥F3 donc minimum 3x2= 6 essais

N ≥ PPCM (6,8,12,16) = 48 ; il faut donc réaliser 48 essais pour avoir l'orthogonalité des facteurs 2 à 2.

```
library(DoE.base)
planort <- oa.design(nlevels=c(2,3,4,4))
planort
   A B C D
1  2 3 1 3
2  1 3 2 4
3  1 2 4 4
4  2 1 1 4
5  2 3 4 2
6  2 2 2 1
7  1 1 2 1
8  2 1 3 1
9  2 2 4 1
10 1 1 3 3
13 2 1 2 4
14 1 1 1 1
15 1 2 4 1
16 1 3 3 1
17 2 3 3 1
18 1 1 3 4
19 2 1 4 1
20 1 2 3 3
21 2 3 2 1
22 2 1 4 2
23 1 3 1 1
26 1 2 3 2
27 1 1 4 3
28 1 2 2 3
29 2 1 2 3
30 2 2 3 2
31 2 3 2 4
32 1 3 4 4
33 2 2 1 2
34 2 2 4 3
35 1 3 2 2
36 1 2 1 1
39 1 3 1 3
40 2 2 2 3
41 1 2 2 2
42 2 2 3 4
43 2 1 3 2
44 2 1 1 3
45 2 3 4 3
46 1 1 4 4
47 1 3 4 2
48 2 3 3 4
class=design, type= oa
```

```
 11  1  3  3  3        24  2  2  1  4        37  2  3  1  2
 12  1  2  1  4        25  1  1  1  2        38  1  1  2  2
```

3. **La jardinerie souhaite mobiliser 4 de ses serres pour cette expérience. Comment procédez-vous pour attribuer les essais à chaque serre ?**

Il est possible d'ajouter un 5ème facteur F5 à 4 modalités avec la fonction ***oa.design*** par exemple :

```
Planort2 <-oa.design(nlevels=c(2,3,4,4,4))
Planort2
    A B C D E     14 1 2 2 4 1     28 2 3 3 4 2     42 1 3 1 2 4
1   2 3 4 1 2     15 2 2 1 4 3     29 2 1 4 3 1     43 1 1 1 4 3
2   1 1 2 2 2     16 1 1 1 1 1     30 2 2 4 2 3     44 1 3 2 2 2
3   1 3 1 1 1     17 1 3 4 4 4     31 1 3 3 4 2     45 2 3 1 2 4
4   2 3 2 1 3     18 2 2 2 4 1     32 1 1 4 1 2     46 2 3 3 2 1
5   1 2 3 1 4     19 2 1 2 1 3     33 2 1 3 1 4     47 2 1 3 4 2
6   2 1 4 2 3     20 1 3 4 3 1     34 2 1 2 4 1     48 1 2 4 4 4
7   2 2 3 2 1     21 2 2 2 3 4     35 1 1 2 3 4      class=design, type= oa
8   2 2 1 3 2     22 2 3 1 4 3     36 2 3 4 3 1
9   1 2 1 1 1     23 2 2 4 1 2     37 1 3 2 1 3
10  1 1 3 3 3     24 1 2 4 2 3     38 1 2 1 3 2
11  1 1 3 2 1     25 2 1 1 2 4     39 2 1 1 3 2
12  1 1 4 4 4     26 1 2 2 2 2     40 2 2 3 1 4
13  2 3 2 3 4     27 1 2 3 3 3     41 1 3 3 3 3
```

4. **Précisez le nombre de degrés de liberté de chaque facteur. La jardinerie souhaiterait également étudier l'effet de la variété du plant de tomate en testant 36 variétés. Serait-il possible d'estimer cet effet variété ? Que pensez-vous des confusions de l'effet variété avec les autres facteurs ?**

48 essais, 3 facteurs à 4 modalités (F1, F4 , F5), 1 facteur à 3 modalités (F3), la constante, il reste alors 35 ddl :
48 – ((4-1) + (4-1) + (4-1) + (3-1) + (2-1) + 1) = 48 – 13 = 35 ddl
Pour introduire l'effet variété, il faudrait introduire un 6ème facteur à 36 modalités. Avec 48 essais réalisés, il reste 35 degrés de liberté ce qui est suffisant pour estimer ce facteur. Cependant avec seulement 48 essais, certaines variétés ne seront testées qu'une seule fois et d'autres deux fois. Il y aura ainsi d'énormes confusions entre l'effet variété et chacun des autres effets (par exemple une variété testée une fois ne sera testée qu'à une seule température).

Exercice 2. Conditions de germination d'un lot de graines de colza (exemple réel, par Alexandra Candeille)

En agronomie, le taux de germination d'un lot de grains peut être un paramètre important pour l'élaboration du rendement. On sait que la germination d'une graine dépend de plusieurs facteurs : humidité, conditions de lumière, de température, substrat…

Dans cette première partie, on suppose que la germination d'une graine de colza dépend de 3 facteurs qualitatifs principaux :

- La température du milieu : chaud ou froid
- La luminosité : sombre ou lumineux
- Le substrat : coton, terreau, terre ou compost

La variable réponse Y correspond au pourcentage de graines qui ont germé au bout de 10 jours d'expérience.

1. **Donner le plan complet de l'expérience.**

On peut assimiler le facteur à 4 modalités comme deux facteurs à 2 modalités. Ainsi, on assimile ce plan asymétrique 4×2^2 à un plan fractionnaire 2^4. Le plan complet comporte donc 16 essais. On peut facilement construire un tel plan avec la fonction `fac.design`.

```
library(DoE.base)
fac.design(nlevels=c(4,2,2),
           factor.names=list(substrat=c("coton","terreau","terre","compost"),
                  lum=c("lumiere","obscurite"),temperature=c("chaud","froid")))

creating full factorial with 16 runs ...

   substrat       lum temperature
1     coton   lumiere       chaud
2   terreau   lumiere       chaud
3     coton obscurite       froid
4     coton obscurite       chaud
5   compost obscurite       chaud
```

```
6   compost   lumiere      froid
7     terre obscurite      chaud
8   compost   lumiere      chaud
9     terre   lumiere      froid
10  terreau obscurite      froid
11  terreau obscurite      chaud
12  compost obscurite      froid
13    terre obscurite      froid
14    terre   lumiere      chaud
15  terreau   lumiere      froid
16    coton   lumiere      froid
class=design, type= full factorial
```

On se limite à 8 essais. Quel plan allez-vous réaliser ? Pourra-t-on intégrer les effets interactions dans le modèle ?

Pour construire ce plan, je suppose que la variable substrat à 4 modalités équivaut à 2 variables à 2 modalités. Je construits donc un plan fractionnaire avec la fonction FrF2.

```
library(FrF2)
plan <- FrF2(nruns=8, nfactors=4, random=FALSE,
         factor.names = list(lum=c("lumiere","obscurite"),
         substrat1=c("-1","1"),  substrat2=c("-1","1"), temperature=c("chaud,froid")))
plan
```

On obtient donc le plan fractionnaire 2^{4-1} :

```
        lum substrat1 substrat2 temperature
1   lumiere        -1        -1       chaud
2 obscurite        -1        -1       froid
3   lumiere         1        -1       froid
4 obscurite         1        -1       chaud
5   lumiere        -1         1       froid
6 obscurite        -1         1       chaud
7   lumiere         1         1       chaud
8 obscurite         1         1       froid
class=design, type= FrF2
```

Il suffit ensuite de combiner manuellement les variables substrat1 et substrat2. Prenons par exemple :

Substrat1	Substrat2		Substrat
-1	-1		Coton
1	-1	→	Terreau
-1	1		Terre
1	1		Compost

Cela nous donne alors le plan fractionnaire suivant :

Essai	*Lumière*	*Substrat*	*Température*
1	Lumière	Coton	Chaud
2	Obscurité	Coton	Froid
3	Lumière	Terreau	Froid
4	Obscurité	Terreau	Chaud
5	Lumière	Terre	Froid
6	Obscurité	Terre	Chaud
7	Lumière	Compost	Chaud
8	Obscurité	Compost	Froid

Pour étudier le modèle complet avec toutes les interactions de degré 2,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{kj} + \varepsilon_{ijk}$$

il faudrait estimer la constante (1), deux facteurs à deux modalités ($2 \times 1 = 2$), un facteur à 4 modalités (3) et les interactions de degré 2 ($2 \times 2 + 2 \times 4 + 2 \times 4 = 20$) ce qui fait au total 26 paramètres à estimer. Or, on n'a que 8 essais dans notre plan d'expérience, donc **on ne pourra pas estimer les interactions**. On choisit alors le modèle d'analyse de la variance suivant :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$

α_i= l'effet du substrat

β_j = l'effet de la luminosité

γ_k = l'effet de la température

ε_{ijk} = la résiduelle

Si l'on n'étudie pas les interactions, le nombre de paramètres à estimer est alors :

$$ddl = 1 + 2 \times (2-1) + (4-1) = 1 + 2 + 3 = 6$$

2. Commentez la qualité du plan.

```
plan_manuel <- data.frame(matrix(c("lumiere", "obscur", "lumiere", "obscur", "lumiere","obscur",
"lumiere", "obscur", "coton", "coton", "terreau", "terreau", "terre", "terre", "compost",
"compost", "chaud","froid","froid","chaud","froid", "chaud","chaud","froid"), nrow=8,ncol=3))
names(plan_manuel) <- c("lum", "sub", "temp")
x <- model.matrix(~.,data=plan_manuel)
print(round(solve(t(x)%*%x),2))
```

```
            (Intercept) lumobscur subcoton subterre subterreau tempfroid
(Intercept)        0.75     -0.25     -0.5     -0.5       -0.5     -0.25
lumobscur         -0.25      0.50      0.0      0.0        0.0      0.00
subcoton          -0.50      0.00      1.0      0.5        0.5      0.00
subterre          -0.50      0.00      0.5      1.0        0.5      0.00
subterreau        -0.50      0.00      0.5      0.5        1.0      0.00
tempfroid         -0.25      0.00      0.0      0.0        0.0      0.50
```

On voit que tous les facteurs sont confondus avec la constante, donc le plan n'est pas optimal. Mais les confusions sont toujours plus faibles que les coefficients diagonaux, bien qu'elles soient quand même élevées. En revanche, les autres coefficients diagonaux sont tous nuls, donc il n'y a pas de confusion entre les facteurs, ce qui est satisfaisant.

3. Réaliser l'expérience et commenter la qualité du modèle choisi.

On décide de s'intéresser à deux paramètres étudiés : le taux de germination et la longueur moyenne du germe après 13 jours l'expérience. Voici les résultats obtenus *(données recueillies réelles) :*

Essai	*Lumière*	*Substrat*	*Température*	*Pourcentage de graines germées le 13e jour*	*Longueur moyenne du germe le 13e jour (mm)*
1	Lumière	Coton	Chaud	0	0
2	Obscurité	Coton	Froid	20	0.5
3	Lumière	Terreau	Froid	30	2.5
4	Obscurité	Terreau	Chaud	35	2.5
5	Lumière	Terre	Froid	25	0.5
6	Obscurité	Terre	Chaud	20	0.5
7	Lumière	Compost	Chaud	30	0.5
8	Obscurité	Compost	Froid	25	7

```
dta<- data.frame(matrix(c("lumiere","obscur","lumiere","obscur","lumiere","obscur","lumiere",
"obscur","coton","coton","terreau","terreau","terre","terre","compost", "compost",
"chaud","froid","froid","chaud","froid","chaud","chaud","froid",
0,20,30,35,25,20,30,25, 0,0.5,2.5,2.5,0.5,0.5,0.5,7) ,nrow=8,ncol=5))
names(dta)<-c("lum","sub","temp","pourcent","longueur")
taux_germi<-lm(pourcent~lum+sub+temp,data=dta)
anova(taux_germi)
Analysis of Variance Table

Response: pourcent
          Df Sum Sq Mean Sq F value Pr(>F)
lum        1  28.13  28.125  0.3103 0.6335
sub        3 559.38 186.458  2.0575 0.3436
temp       1  28.13  28.125  0.3103 0.6335
Residuals  2 181.25  90.625

taille_germe<-lm(longueur~lum+sub+temp,data=dta)
anova(taille_germe)
```

```
Analysis of Variance Table

Response: longueur
          Df Sum Sq Mean Sq F value Pr(>F)
lum        1  6.125  6.1250  1.3611 0.3636
sub        3 16.750  5.5833  1.2407 0.4754
temp       1  6.125  6.1250  1.3611 0.3636
Residuals  2  9.000  4.5000
```

Malheureusement, les résultats de l'analyse de variance montrent qu'avec notre expérience, aucun des facteurs (substrat, luminosité et température) n'impacte significativement ni le taux de germination des graines, ni la taille moyenne du germe.

On peut envisager qu'avec un plan d'expérience complet (16 essais au lieu de 8), on aurait eu des résultats différents. En effet, l'expérience a présenté plusieurs limites :

- On n'a utilisé que 20 graines par lot (donc 20 répétitions), mais on peut imaginer en faire germer plus.
- L'expérience a duré 13 jours, car on était limité dans le temps, mais dans des conditions optimales on aurait pu attendre plus longtemps pour laisser aux graines le temps de germer.

Exercice 3. Caractérisation de la texture de mousse de blanc d'œufs (exemple réel, par Enora Delourme)

Une entreprise d'ovoproduits souhaite caractériser la texture des mousses de blanc d'œufs. Afin de minimiser les paramètres à étudier, 4 facteurs qui pourraient influencer la texture des mousses de blancs d'œufs sont sélectionnés. Ce sont les suivants :

- La température de pasteurisation (F1) : 57, 59 ou 60
- La stabilité de la mousse (F2) : faible moyenne haute
- La viscosité du blanc d'œuf (F3) : faible moyenne haute
- Le type d'élevage (F4) : sol, plein air, cage ou bio

La texture des mousses se mesure à l'aide d'un texturomètre mesurant la force des mousses (exprimé en N) qui constitue la variable Y.

1. Combien d'essais sont nécessaires pour obtenir un plan où les facteurs sont orthogonaux 2 à 2 ?

Nous avons 2 facteurs à 2 modalités, 1 facteur à 3 modalités et 1 facteur à 4 modalités, ce qui nous donne :

- Un nombre d'essais maximum de : 3^3 x 4 = 108 essais
- Un nombre d'essais minimal dépendant du nombre de paramètres à estimer :
  1 + 3 x (3-1) + (4-1) = 10 paramètres
- Pour l'orthogonalité de F1 et F2 ; F1 et F3 ; F1 et F4 = 4 x 3 = 12 essais
- Pour l'orthogonalité de F2 et F3 ; F2 et F4 ; F4 et F3 = 3 x 3 = 9 essais
- ⇨ Le PPCM de 12 et 9 étant 36, le seul plan complet permettant d'obtenir des facteurs orthogonaux 2 à 2 est un plan à 36 essais.

2. On souhaite réaliser 16 essais maximum. Est-ce possible de construire un plan en 16 essais ?

Oui, il est possible de construire un plan en 16 essais car tous les paramètres seraient bien estimés et il resterait même 6 DDL pour la résiduelle. Cependant, la qualité du plan sera à vérifier.

3. Construire un plan $L_{16}3^34$ à la main, par la méthode des carrés latins mutuellement orthogonaux, en détaillant les étapes de construction :

1- On commence par construire deux carrés latins de taille 4 (3 facteurs à 4 modalités) avec :

F1 en jaune ; F2 en bleu, F3 en rouge et F4 en vert

	1	2	3	4
1	1	2	3	4
2	2	3	4	1
3	3	4	1	2
4	4	1	2	3

	1	2	3	4
1	1	2	3	4
2	3	4	1	2
3	4	3	2	1
4	2	1	4	3

2- Ensuite on déplie les plans

F1	F2	F3	F4
1	1	1	1
1	2	2	2
1	3	3	3
1	4	4	4
2	1	2	3
2	2	3	4
2	3	4	1
2	4	1	2
3	1	3	4
3	2	4	3
3	3	1	2
3	4	2	1
4	1	4	2
4	2	1	1
4	3	2	4
4	4	3	3

3- Enfin, pour les facteurs à 3 modalités, on va compresser 2 niveaux ensemble :

F1	F2	F3	F4
1	1	1	1
1	2	2	2
1	2	2	3
1	3	3	4
2	1	2	3
2	2	2	4
2	2	3	1
2	3	1	2
2	1	2	4
2	2	3	3
2	2	1	2
2	3	2	1
3	1	3	2
3	2	1	1
3	2	2	4
3	3	2	3

On obtient donc un plan en 16 essais avec 3 facteurs à 3 modalités et 1 facteur à 4 modalités.

4. Vérifier la qualité du plan de la matrice des essais sur R.

Après importation du jeu de la matrice des effets, on analyse la qualité du plan par les lignes de code :

```
plan2 <- read.table('MOLS-exo2.csv',sep=';',header=TRUE)
for (i in 1:4) plan2[,i] <- as.factor(plan2[,i])
summary(plan2)
X <- model.matrix(~.,data=plan2)
> solve(t(X)%*%X)
            (Intercept)    F12   F13    F22   F23           F32   F33           F42    F43    F44
(Intercept)       0.750 -0.250 -0.25 -0.250 -0.25 -2.500000e-01 -0.25 -2.500000e-01 -0.125 -0.125
F12              -0.250  0.375  0.25  0.000  0.00  0.000000e+00  0.00  0.000000e+00  0.000  0.000
F13              -0.250  0.250  0.50  0.000  0.00  0.000000e+00  0.00  0.000000e+00  0.000  0.000
F22              -0.250  0.000  0.00  0.375  0.25  0.000000e+00  0.00  0.000000e+00  0.000  0.000
F23              -0.250  0.000  0.00  0.250  0.50  0.000000e+00  0.00  0.000000e+00  0.000  0.000
F32              -0.250  0.000  0.00  0.000  0.00  6.000000e-01  0.40  4.163336e-17 -0.300 -0.300
F33              -0.250  0.000  0.00  0.000  0.00  4.000000e-01  0.60  2.775558e-17 -0.200 -0.200
F42              -0.250  0.000  0.00  0.000  0.00  2.114711e-17  0.00  5.000000e-01  0.250  0.250
F43              -0.125  0.000  0.00  0.000  0.00 -3.000000e-01 -0.20  2.500000e-01  0.650  0.400
F44              -0.125  0.000  0.00  0.000  0.00 -3.000000e-01 -0.20  2.500000e-01  0.400  0.650
```

On obtient la matrice des essais $(X'X)^{-1}$. On observe que la matrice est bien diagonale, les effets sont un peu confondus avec la constante (mais les confusions sont assez faibles). Dans la matrice, il y a quelques confusions mais elles sont relativement faibles donc nous pouvons donc dire que ce plan est tout de même de bonne qualité.

5. **Les essais ont été réalisés et les réponses associées à chaque essai sont présentés dans le tableau suivant. Analyser les résultats par analyse de variance et conclure sur les facteurs qui influence la texture des mousses de blanc.**

T° pasteurisation (F1)	Stabilité (F2)	Viscosité (F3)	Type d'élevage (F4)	Force mousse (Y)
57	F	F	sol	78,5
57	M	M	pa	68,8
57	M	M	cage	65,6
57	H	H	bio	49,6
59	F	M	cage	95,1
59	M	M	bio	71,2
59	M	H	sol	86,7
59	H	F	pa	44,4
59	F	M	bio	91,5
59	M	H	cage	85,3
59	M	F	pa	61,3
59	H	M	sol	50,2
60	F	H	pa	70,4
60	M	F	sol	61,4
60	M	M	bio	66,7
60	H	M	cage	49,8

Pour analyser les résultats, on réalise une ANOVA où les interactions ne sont pas prises en compte car nous ne pouvons pas estimer les interactions avec ce plan construit à partir de carrés latins.

Avec la ligne de code suivante, on obtient l'analyse de variance pour notre modèle avec la sortie associée :

```
AovSum(Y~F1+F2+F3+F4, data = plan2)
```

```
Ftest
             SS df      MS F value    Pr(>F)
F1        25.21  1   25.21  0.3085 0.5897146
F2      2502.78  1 2502.78 30.6339 0.0001767 ***
F3       229.50  1  229.50  2.8091 0.1218887
F4         1.71  1    1.71  0.0209 0.8875449
Residuals 898.70 11   81.70
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ttest
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 95.39375   11.74173  8.1243    1e-05 ***
F1          -1.77500    3.19569 -0.5554  0.58971
F2         -17.68750    3.19569 -5.5348  0.00018 ***
F3           5.64583    3.36856  1.6760  0.12189
F4           0.30833    2.13046  0.1447  0.88754
```

D'après l'analyse de variance, il semble que seule la stabilité du blanc d'œuf (F2) ait une probabilité critique inférieure à 5%. Le facteur stabilité a donc un effet significatif sur la texture des mousses de blancs d'œufs.

Exercice 4. Optimisation de la production de levures Saccharomyces cerevisiae (exemple réel, par Claire Morice)

Une entreprise de production de levure *Saccharomyces cerevisiae* cherche à optimiser ses rendements, en déterminant les conditions les plus adaptées au développement de cette levure. Cette levure d'intérêt est utilisée en agroalimentaire pour la fermentation de produits comme le pain ou les boissons alcoolisées.

La variable réponse à étudier est la production de levure *Saccharomyces cerevisiae* en g/batch

Le chef de production a réussi à identifier les paramètres qui influencent le développement de cette levure, et souhaite les faire varier de la façon suivante :

Paramètres :

- F1 : Température (°C) : 35, 50
- F2 : pH : 5.5, 6.5

- F3 : durée d'incubation (heure) : 2, 4
- F4 : Teneur en glucose (g/L) : 30, 32, 34, 36

Question : Contraint par le temps, le chef de production n'a pas d'autres choix que de réaliser les 16 essais dans 2 bioréacteurs de tailles différentes (8 essais par bioréacteur). Proposez au chef de production, un plan d'expérience prenant en compte cette contrainte. Avant de commencer, donnez le nom du plan à construire. Ensuite, donnez la matrice des effets et expliquez-lui en détail les essais à réaliser.

Nom du plan à construire : $L_{16}4^1 2^4$

Pour prendre en compte l'effet de la taille du bioréacteur, on ajoute un 5$^{\text{ème}}$ facteur à 2 modalités.

On sait qu'ajouter un facteur à 4 modalités peut revenir à ajouter 2 facteurs à 2 modalités. On construit d'abord la matrice des effets pour les facteurs Température, pH, durée et bioréacteur.

- F5 : taille du bioréacteur : moyen, grand

I	F1	F2	F3	F5	F12	F13	F15	F23	F25	F35	F123	F125	F135	F235	F1235
1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
1	+1	+1	+1	-1	+1	+1	-1	+1	-1	-1	+1	-1	-1	-1	-1
1	+1	+1	-1	+1	+1	-1	+1	-1	+1	-1	-1	+1	-1	-1	-1
1	+1	+1	-1	-1	+1	-1	-1	-1	-1	+1	-1	-1	+1	+1	+1
1	+1	-1	+1	+1	-1	+1	+1	-1	-1	+1	-1	-1	+1	-1	-1
1	+1	-1	+1	-1	-1	+1	-1	-1	+1	-1	-1	+1	-1	+1	+1
1	+1	-1	-1	+1	-1	-1	+1	+1	-1	-1	+1	-1	-1	+1	+1
1	+1	-1	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1	+1	-1	-1
1	-1	+1	+1	+1	-1	+1	-1	+1	+1	+1	-1	-1	+1	+1	-1
1	-1	+1	+1	-1	-1	+1	+1	+1	-1	-1	-1	+1	-1	-1	+1
1	-1	+1	-1	+1	-1	+1	-1	-1	+1	-1	+1	-1	+1	-1	+1
1	-1	+1	-1	-1	-1	+1	+1	-1	-1	+1	+1	+1	-1	+1	-1
1	-1	-1	+1	+1	+1	-1	-1	-1	-1	+1	+1	+1	-1	-1	+1
1	-1	-1	+1	-1	+1	-1	+1	-1	+1	-1	+1	-1	+1	+1	-1
1	-1	-1	-1	+1	+1	+1	-1	+1	-1	-1	-1	+1	+1	+1	-1
1	-1	-1	-1	-1	+1	+1	+1	+1	+1	+1	-1	-1	-1	-1	+1

On doit donc construire un plan avec 4 facteurs à 2 modalités (F1, F2, F3, F5) et 1 facteur à 4 modalités (F4).

Pour ajouter le facteur F4 « glucose », on peut ajouter 2 facteurs à 2 niveaux : A et B. On confond les facteurs A et B avec des interactions d'ordre le plus élevé moins 1: ici on prend F123 et F125

Codage : - - = 1, - + = 2 ; + - = 3 ; + + = 4

F123 (A)	F125 (B)	F4
+1	+1	4
+1	-1	3
-1	+1	2
-1	-1	1
-1	-1	1
-1	+1	2
+1	-1	3
+1	+1	4
-1	-1	1
-1	+1	2
+1	-1	3
+1	+1	4
+1	+1	4
+1	-1	3
-1	+1	2
-1	-1	1

Matrice des essais en valeurs codées :

Essai	F1	F2	F3	F5	F4
1	+1	+1	+1	+1	4
2	+1	+1	+1	-1	3
3	+1	+1	-1	+1	2
4	+1	+1	-1	-1	1
5	+1	-1	+1	+1	1
6	+1	-1	+1	-1	2
7	+1	-1	-1	+1	3
8	+1	-1	-1	-1	4
9	-1	+1	+1	+1	1
10	-1	+1	+1	-1	2
11	-1	+1	-1	+1	3
12	-1	+1	-1	-1	4
13	-1	-1	+1	+1	4
14	-1	-1	+1	-1	3
15	-1	-1	-1	+1	2
16	-1	-1	-1	-1	1

Expériences à réaliser par le chef de production (matrice des essais en valeurs réelles) :

Essai	Température (°C)	pH	Durée (heures)	Bioréacteur	Teneur en glucose (g/L)
1	35	5,5	2	Moyen	36
2	35	5,5	2	Grand	34
3	35	5,5	4	Moyen	32
4	35	5,5	4	Grand	30
5	35	6,5	2	Moyen	30
6	35	6,5	2	Grand	32
7	35	6,5	-1	Moyen	34
8	35	6,5	-1	Grand	36
9	50	5,5	2	Moyen	30
10	50	5,5	2	Grand	32
11	50	5,5	4	Moyen	34
12	50	5,5	4	Grand	36
13	50	6,5	2	Moyen	36
14	50	6,5	2	Grand	34
15	50	6,5	4	Moyen	32
16	50	6,5	4	Grand	30

Exercice 5. Comment fabriquer des jouets (exemple fictif, par Oriane David) ?

Principe de l'exercice inspiré d'exemple étudié en cours à l'université d'Alberta. Les données ont été un peu changées et la mise en contexte et les questions ont été totalement inventées.

Un fabricant de jouet, M. Rapido, souhaite tester l'efficacité de quatre nouvelles méthodes (0,1,2,3) d'assemblage de jouets sur le temps de fabrication des jouets dans son usine. Il fait pour cela appel à quatre opérateurs différents, non formés au préalable à ces méthodes. La fabrication de jouet est très fatigante pour tous les opérateurs. Le fabriquant est très pressé et ne dispose seulement que d' une demie- journée pour faire les essais. Il ne peut donc en faire que 16.

1. **Quels sont les facteurs que l'on cherche à étudier ? Quels sont les facteurs que l'on cherche à contrôler ?**

Le facteur que l'on cherche à étudier est la méthode d'assemblage. On cherche à contrôler l'effet des opérateurs ainsi que l'ordre de passage des essais des différentes méthodes (car la fabrication est fatigante).

2. **Quel est le modèle à étudier ? Pourquoi ? Quels sont les degrés de liberté relatifs au modèle ?**

   Quel type de plan d'expérience peut-on mettre en place ? Justifier.
   Ecrire un plan d'expérience pouvant être mis en place afin d'étudier la vitesse de fabrication des jouets.

Le modèle à étudier est :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$
$$avec\ \varepsilon \rightarrow N(0,1)$$

On ne prend pas les interactions en compte car elles sont supposées nulles :

- Les opérateurs ne connaissent pas les méthodes, on peut donc supposer qu'il n'y a pas d'interaction entre les opérateurs et la méthode
- La fabrication est également fatigante quelle que soit l'opérateur ou la méthode, il n'y a donc a priori pas d'interaction entre l'ordre de passage et les autres facteurs.

Il y a 16 degrés de liberté en tout avec : 3 d.l pour chaque facteur, 1 d.l pour la constante et 6 d.l pour la résiduelle.

On a trois facteurs avec un même nombre de modalités (à 4 modalités chacun) et 16 essais. Il s'agit donc d'un plan de type $L_{16}4^3$. Il s'agit d'un plan de type carré latin.

Opérateur **Ordre**	**0**	**1**	**2**	**3**
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

3. **Le fabricant a entendu parler d'ergonomie au travail. Il veut tester quatre plans de travail différents (0,1,2,3) . Mais il dispose du même nombre d'essais. Ecrire le nouveau modèle adapté ainsi que les degrés de liberté et un nouveau plan d'expérience.**

Le nouveau modèle est

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \sigma_l + \varepsilon_{ijkl}$$
$$avec\ \varepsilon \rightarrow N(0,1)$$

Il y a alors 3 ddl pour chacun des 4 facteurs étudiés, 1 pour la constante et 3 pour la résiduelle.

On nomme le type de plan relatif à ce modèle un carré gréco-latin. On peut utiliser un MOLS du plan précédemment trouvé pour construire ce nouveau plan (utilisation des tables). On peut aussi créer un plan à partir d'un plan 2^{8-4} en codant chaque facteur (à 4 modalités) par 2 facteurs à 2 modalités.

MOLS directement :

Opérateur **Ordre**	**0**	**1**	**2**	**3**
0	00	11	22	33
1	31	22	13	00
2	12	03	30	21
3	23	30	01	12

2^{8-4} :

A	B	C	D	AB	BC	CD	AD	Facteur 1	Facteur 2	Facteur 3 (AB-BC)	Facteur 4 (CD-AD)
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	-1	1	1	-1	-1	1	2	1	4
1	1	-1	1	1	-1	-1	1	1	3	2	3
1	1	-1	-1	1	-1	1	-1	1	4	2	2
1	-1	1	1	-1	-1	1	1	2	1	4	1
1	-1	1	-1	-1	-1	-1	-1	2	2	4	4

1	-1	-1	1	-1	1	-1	1	2	3	3	3
1	-1	-1	-1	-1	1	1	-1	2	4	3	2
-1	1	1	1	-1	1	1	-1	3	1	3	2
-1	1	1	-1	-1	1	-1	1	3	2	3	3
-1	1	-1	1	-1	-1	-1	-1	3	3	4	4
-1	1	-1	-1	-1	-1	1	1	3	4	4	1
-1	-1	1	1	1	-1	1	-1	4	1	2	2
-1	-1	1	-1	1	-1	-1	1	4	2	2	3
-1	-1	-1	1	1	1	-1	-1	4	3	1	4
-1	-1	-1	-1	1	1	1	1	4	4	1	1

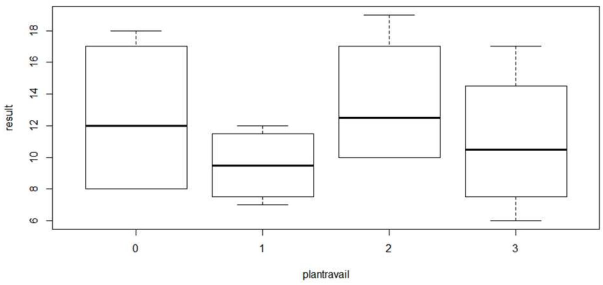
4. **Analyser les durées de fabrication des jouets relatifs aux essais dans le tableau suivant. Quelle décision doit prendre M. Rapido pour optimiser les conditions de travail dans son usine ?**

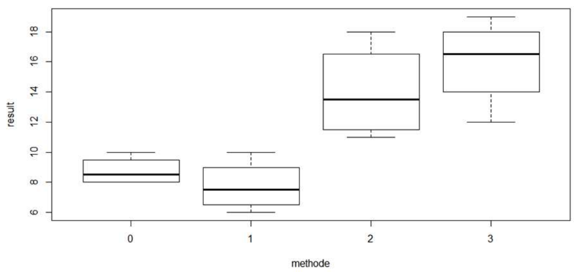
Opérateur / Ordre	0	1	2	3
0	2, 1= 11	1, 2= 10	3, 3=17	0,0=8
1	1, 0=8	2, 3= 12	0, 2=10	3,1=12
2	0,3= 9	3, 0= 16	1, 1=7	2, 2=15
3	3, 2= 19	0, 1=8	2, 0=18	1,3=6

```
operateur= rep (c("0","1","2","3"), 4)
ordre=c("0","0","0","0","1","1","1","1","2","2","2","2","3","3","3","3")
methode=c("2","1", "3", "0", "1", "2", "0", "3", "0", "3", "1", "2", "3", "0", "2", "1")
plantravail=c("1","2", "3", "0", "0","3", "2", "1", "3", "0", "1", "2", "2", "1", "0", "3")
result=c(11, 10, 17, 8, 8, 12,10, 12, 9, 16, 7, 15, 19, 8, 18, 6)
data=data.frame(ordre, operateur, methode, plantravail, result)
summary(data)
mod.lm=lm(result~ordre+operateur+methode+plantravail, data=data)
anova(mod.lm)
Analysis of Variance Table

Response: result
            Df Sum Sq Mean Sq F value   Pr(>F)
ordre        3  10.25   3.417  1.9524 0.298258
operateur    3  15.25   5.083  2.9048 0.202252
methode      3 192.25  64.083 36.6190 0.007299 **
plantravail  3  36.75  12.250  7.0000 0.072147 .
Residuals    3   5.25   1.750
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La méthode utilisée a un effet significatif sur le temps de fabrication du jouet au seuil de 5%. Le plan de travail a un effet marginalement significatif (seuil de 8%). Les facteurs de contrôle (opérateur et ordre) n'ont pas d'effet sur la durée de fabrication du jouet.

```
boxplot(result~plantravail)
boxplot(result~methode)
```





On peut dire que la méthode la plus efficace est la méthode 3 suivie de près par la méthode 2. Il est plus difficile d'avoir une lecture claire concernant le plan de travail. Il serait cependant nécessaire d'étudier la distribution des résidus afin de pouvoir attester que le modèle est adapté. Il serait également nécessaire de mener des test (test de Tukey par exemple) afin de voir quelles méthodes ont des effets significativement différents.

5. **M. Rapido adore la musique (de Noel). Mais il veut vérifier que la présence de celle-ci ne ralentisse pas la fabrication des jouets. Peut- il observer l'effet de la présence de musique en plus de tous les autres en seulement 16 essais ? Justifier.**

Dans ce modèle, il reste 3 degrés de liberté pour la résiduelle. Il reste donc assez de degré de liberté pour étudier le facteur musique avec 2 (sans musique ou avec) , 3 (musique forte, pas forte, ou pas de musique) ou 4 modalités (avec 4 intensité sonores). Dans le dernier cas, on ne pourrait pas avoir de résiduelle ce qui rendrait plus difficile l'étude des résultats. De plus l'existence d'un troisième MOLS dans les tables, prouve que jusqu'à 5 facteurs à 4 modalités peuvent être étudiés.

Exercice 6. Silence, ça pousse ! (exemple fictif, par Tom AYRAULT)

Un étudiant d'Agrocampus Ouest est confiné chez lui à cause de l'épidémie de Covid-19. Dans le cadre d'un cours d'écologie, il doit discuter des différents effets qui influent sur la vitesse de croissance d'une plante. Il finit par mettre en évidence 3 facteurs principaux : la luminosité, la quantité d'engrais, la quantité d'eau.
Très motivé, il décide de vérifier l'importance de ces facteurs en plantant lui-même des graines de haricot chez lui en jouant sur les paramètres évoqués.
On obtient ainsi les trois facteurs à 3 modalités chacun :

- La quantité d'engrais (avec engrais, sans engrais) ➔ facteur F1 à 3 modalités
- La luminosité (5h, 10h ou 15h par jour) ➔ facteur F3 à 3 modalités
- La quantité d'eau (peu, moyen, beaucoup) ➔ facteur F4 à 3 modalités

L'étudiant ne dispose que de 10 pots. Il faut créer un plan d'expérience.

1. **Quel type de plan mettriez-vous en place ? Dessinez-le. Quelle conséquence pour l'étude des interactions ?**

En présence de 3 facteurs à 3 modalités et ne pouvant excéder 10 essais, on peut proposer l'utilisation d'un **carré latin**. L'utilisation du carré latin permet d'étudier en 9 essais les trois effets principaux à 3 modalités.
Cependant, il ne reste plus de degré de liberté pour étudier les effets d'interactions, ce qui aurait été utile à l'analyse.

F1 \ F2	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

F3

Carré latin

	F1 quantité d'engrais	F2 luminosité	F3 quantité d'eau
1er essai	0	0	0
2ème essai	0	1	1
3ème essai	0	2	2
4ème essai	1	0	1
5ème essai	1	1	2
6ème essai	1	2	0
7ème essai	2	0	2
8ème essai	2	1	0
9ème essai	2	2	1

2. **Comment aurait-on pu étudier des effets d'interactions ?**

Il aurait fallu augmenter les essais et ainsi augmenter le nombre de degré de liberté à notre disposition.

3. **Recopiez le tableau sur Excel. Evaluez la qualité du plan obtenu sur R.**

```
latin <- read.csv("latin.csv",sep=";")
View(latin)
X=model.matrix(~.,data=latin)
solve(t(X)%*%X)
            (Intercept)         F1         F2         F3
(Intercept)   0.6111111 -0.1666667 -0.1666667 -0.1666667
F1           -0.1666667  0.1666667  0.0000000  0.0000000
F2           -0.1666667  0.0000000  0.1666667  0.0000000
F3           -0.1666667  0.0000000  0.0000000  0.1666667
```

La matrice $(XX')^{-1}$ est diagonale, les effets principaux sont alors orthogonaux. C'est en effet une propriété du carré latin.

Avec un plan d'expérience en poche, l'étudiant va dans son garage chercher les pots. Il découvre qu'il possède finalement une vingtaine de pots. Avec ces nouveaux degrés de liberté à sa disposition, il se dit qu'il peut tester un nouveau facteur F4, le facteur terre. En effet il possède un sac de terreau et de la terre de son potager. Etant difficile de mélanger avec précision des terres, ce facteur ne comporterait que deux modalités : terreau ou terre du potager.

4. **Quel plan d'expérience proposez-vous ? Dessinez-le.**

Il s'agit ici de construire un **plan avec des nombres de modalités différentes, ici $L_9 2^1 3^3$**. Il faut alors combiner le carré latin précédent à un simple plan 2^1.
Avec 18 essais à disposition, on double le carré latin et on attribue une modalité du facteur F4 à chaque carré latin.

F1 quantité d'engrais	F2 luminosité	F3 quantité d'eau
0	0	0
0	1	1
0	2	2
1	0	1
1	1	2
1	2	0
2	0	2
2	1	0
2	2	1

F4 terre
0
1

0	0	0	0
0	0	1	1
0	0	2	2
0	1	0	1
0	1	1	2
0	1	2	0
0	2	0	2
0	2	1	0
0	2	2	1
1	0	0	0
1	0	1	1
1	0	2	2
1	1	0	1
1	1	1	2
1	1	2	0
1	2	0	2
1	2	1	0
1	2	2	1

5. **De la même manière, évaluez la qualité du plan obtenu sur R.**

```
latin <- read.csv(".csv",sep=";")
View(latin)
X=model.matrix(~.,data=latin)
solve(t(X)%*%X)
             (Intercept)          F1          F2          F3         F4
(Intercept)   0.36111111 -0.08333333 -0.08333333 -0.08333333 -0.1111111
F1           -0.08333333  0.08333333  0.00000000  0.00000000  0.0000000
F2           -0.08333333  0.00000000  0.08333333  0.00000000  0.0000000
F3           -0.08333333  0.00000000  0.00000000  0.08333333  0.0000000
F4           -0.11111111  0.00000000  0.00000000  0.00000000  0.2222222
```

De même on observe que matrice $(XX')^{-1}$ est diagonale, les effets principaux sont alors aussi orthogonaux. Les effets principaux sont alors indépendants. Le facteur terre joue le rôle d'effet bloc.

Exercice 7. Un peu de musique (exemple fictif, par Clarisse Thiard)

On s'intéresse à l'évaluation d'une audition musicale par différents juges (J). Pour cela, ils notent l'interprétation de différents morceaux (MO) joués par différents musiciens (MU) et donnent une note globale d'exécution. L'évaluation est donnée par une note Y

On comptabilise 3 juges, 3 morceaux différents et 3 musiciens différents.

1. **Réaliser le plan d'expérience manuellement pour cette expérience.**

On réalise un carré latin avec 3 facteurs comportant chacun 3 catégories. On a donc J=3.

On obtient ainsi avec les juges en bleu et les musiciens en bleu

	1	2	3
1	1	2	3
2	2	3	1

3	3	1	2

2. **Expliciter le modèle et les différents degrés de liberté du modèle.**

Le modèle s'exprime :

$Y_{ijk} = \mu + J_i + MU_j + MO_k + \varepsilon_{ijk}$

La constante possède un ddl de 1.

Chaque facteur possède 3 modalités donc un ddl de 2.

Le total de ddl et de 9.

On obtient donc un ddl pour la résiduelle de 9-7=2.

3. **On réalise maintenant une 2e session avec 2 juges différents, 2 morceaux différents et 2 musiciens différents. Réaliser le nouveau plan en ajoutant les deux sessions indépendamment.**

On construit un carré latin pour la 2e session.

	1	*2*
1	*1*	*2*
2	*2*	*1*

On obtient ainsi le tableau

1	*1*	*1*
1	*2*	*2*
2	*1*	*2*
2	*2*	*1*

On ajoute donc ce tableau à chaque expérience de la première session. Pour chaque expérience réalisée initialement une fois lors de la première session, on en réalise maintenant 4 avec les modalités de la 2e session.

Exercices sur les plans optimaux

Exercice 1. Optimisation poids de melons (exemple fictif, par Anna Letemplier)

Jeu de données trouvé sur Internet. Données probablement réelles mais triées et simplifiées pour certains facteurs.

Un agriculteur bio vend des melons et souhaite faire une étude pour avoir les plus gros melons possibles. Il se demande quels sont les effets, sur la taille de ses melons, de la couverture, de la date de plantation, de l'apport en matière organique par du fumier, du temps passé en terre et de la variété.

Nous avons donc 5 facteurs pour une variable réponse quantitative, correspondant au poids du melon en gramme.

Facteur 1 Variété	Qualitative	5 modalités	Hugo, Cezanne, Manta, Escrito, Fidji
Facteur 2 Créneau de semis	Qualitative	3 modalités	Précoce Pleine saison Tardif
Facteur 3 Couverture	Qualitative	2 modalités	Couvert Non couvert
Facteur 4 Durée en terre	Quantitative		57 jrs à 91 jrs
Facteur 5 Apports organiques	Qualitative	2 modalités	OUI NON

1. **Quelle serait votre approche face à cette expérience et votre stratégie de planification ?**

On a un plan mixte avec 4 facteurs qualitatifs, et 1 facteur quantitatif.

Deux stratégies sont possibles :

1- Construire un plan avec les variables quantitatives seules et un plan avec les variables qualitatives seules et répéter le plan pour chaque expérience du plan qualitatif.
2- Transformer les variables quantitatives en variables qualitatives avec au moins 3 niveaux (afin de prendre en compte les effets quadratiques) et construire un plan pour les variables qualitatives

Stratégie 1 : Pour le plan avec la variable quantitative, il faut faire quelques essais avec la valeur minimale, quelques essais avec la valeur maximale et quelques essais avec la valeur médiane. Je dois donc faire au minimum 3 essais pour ce plan pour tester ces valeurs au moins une fois.

Pour le plan pour les variables qualitatives (2 facteurs à 2 modalités, 1 facteur à 3 modalités, 1 facteur à 5 modalités), il faut donc PPCM(4,6,10,15) = 60 essais minimum pour avoir l'orthogonalité des facteurs.

⇨ Cela représente un total minimum de 3*60 = 180 essais à réaliser.

Avec une seule variable quantitative il vaut mieux opter pour 2nd stratégie : on transforme la variable quantitative « Durée en terre » en variable qualitative à 3 modalités (57jrs, 73jrs et 91jrs), qui seront représentées par 1,2 ou 3.

2. **Quel plan mettez-vous en place pour les variables quantitatives ? Combien d'essais sont nécessaires ?**

On réalise un plan asymétrique avec des facteurs à 2, 3 et 5 niveaux.

Nombre maximal d'essais : 2^2*3^2*5 = 4*9*5= 180 essais

Nombre minimal d'essais = 1 + 2*(2-1) + 2*(3-1) + (5-1) = 11 essais

Pour l'orthogonalités des facteurs, il faut PPCM(4,6,8,10,15) = 180 essais.

Donc seul le plan complet permet d'avoir l'orthogonalité des facteurs.

3. **Donnez la matrice des essais.**

```
#Je crée plan complet pour 5 facteurs, PPCM(4,6,9,12,15)=180 essais
plan<-gen.factorial(levels=c(2,2,3,3,5),nVars=5, factors = "all", varNames =
c("Couvert","ApportOrg","Tps","Semis","Variete"))
```

4. Évaluez la qualité du plan. Que pouvez-vous en conclure ?

```
#Vérification de la qualité du plan
Options(conrasts=c("contr.sum","contr.sum")) # pour avoir la contrainte somme des alpha_i égale
à 0
X<-model.matrix(~.,data=plan)
XX<-t(X)%*%X
round(solve(XX),3)
```

	(Intercept)	Couvert1	ApportOrg1	Tps1	Tps2	Semis1	Semis2	Variete1	Variete2	Variete3	Variete4
(Intercept)	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Couvert1	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ApportOrg1	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tps1	0.000	0.000	0.000	0.011	-0.006	0.000	0.000	0.000	0.000	0.000	0.000
Tps2	0.000	0.000	0.000	-0.006	0.011	0.000	0.000	0.000	0.000	0.000	0.000
Semis1	0.000	0.000	0.000	0.000	0.000	0.011	-0.006	0.000	0.000	0.000	0.000
Semis2	0.000	0.000	0.000	0.000	0.000	-0.006	0.011	0.000	0.000	0.000	0.000
Variete1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.022	-0.006	-0.006	-0.006
Variete2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.006	0.022	-0.006	-0.006
Variete3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.006	-0.006	0.022	-0.006
Variete4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.006	-0.006	-0.006	0.022

Le plan est de très bonne qualité ; les effets sont orthogonaux entre eux et il y a très peu de confusions.

Cependant le nombre d'essais est beaucoup trop élevé ; cela demande des investissements très important et ce n'est donc pas toujours faisable.

5. L'agriculteur nous informe qu'il peut faire seulement 40 essais. Quelle est votre nouvelle approche ?

On ne change pas de plan ; on reste sur un plan asymétrique à 5 facteurs (2, 3 et 5 niveaux).

Cependant, on utilise l'algorithme de Federov pour trouver le plan optimal. Cela détériore forcément la qualité de notre plan ; il faut donc vérifier la qualité du plan obtenu.

6. Donnez la nouvelle matrice des essais.

Voici les codes permettant d'obtenir cette matrice sur R :

```
#Optimisation du plan en 40 essais
planopt<-optFederov(~.,data=plan,nTrials = 40, nRepeats = 100, criterion="D")
planopt$design
#j'etudie les facteurs en 40 essais et repete 100 fois pour eviter un minimum local

#Nouvelle matrice des effets
X<-model.matrix(~.,data=planopt$design)
```

Cette matrice nous permet de réduire à 40 le nombre d'essais.

7. Vérifiez la qualité de ce nouveau plan. Comparez cette dernière avec celle du plan complet.

```
#Vérification de la qualité du nouveau plan
round(solve(t(X)%*%X),3)
```

	(Intercept)	Couvert1	ApportOrg1	Tps1	Tps2	Semis1	Semis2	Variete1	Variete2	Variete3	Variete4
(Intercept)	0.025	0.000	0.000	-0.001	0.003	-0.001	0.003	0.000	0.000	0.000	-0.001
Couvert1	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ApportOrg1	0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tps1	-0.001	0.000	0.000	0.050	-0.027	0.000	0.001	-0.002	-0.002	-0.002	0.007
Tps2	0.003	0.000	0.000	-0.027	0.054	0.001	-0.002	0.004	0.004	0.004	-0.006
Semis1	-0.001	0.000	0.000	0.000	0.001	0.050	-0.027	-0.002	-0.002	-0.002	0.007
Semis2	0.003	0.000	0.000	0.001	-0.002	-0.027	0.054	0.004	0.004	0.004	-0.006
Variete1	0.000	0.000	0.000	-0.002	0.004	-0.002	0.004	0.101	-0.024	-0.024	-0.026
Variete2	0.000	0.000	0.000	-0.002	0.004	-0.002	0.004	-0.024	0.101	-0.024	-0.026
Variete3	0.000	0.000	0.000	-0.002	0.004	-0.002	0.004	-0.024	-0.024	0.101	-0.026
Variete4	-0.001	0.000	0.000	0.007	-0.006	0.007	-0.006	-0.026	-0.026	-0.026	0.102

Malgré le petit nombre d'essais par rapport au plan complet, on garde une bonne orthogonalité entre les facteurs. On a perdu néanmoins en qualité ; on voit apparaitre de légères confusions entre les facteurs entre eux, et avec la constante. Néanmoins les coefficients sont faibles donc ces confusions sont faibles.

```
#Comparaison qualité du plan par rapport au plan complet
planopt$D
[1] 0.5246114
```

On a une efficacité de 50% pour l'estimation de chaque facteur : cette estimation est convenable au vu de l'importante diminution du nombre d'essais.

Ce plan correspond à la meilleure option pour cet agriculteur qui ne veut pas réaliser plus de 40 essais.

Il serait intéressant d'évaluer la qualité du plan en baissant d'avantage le nombre d'essais, le but étant de trouver le meilleur compromis.

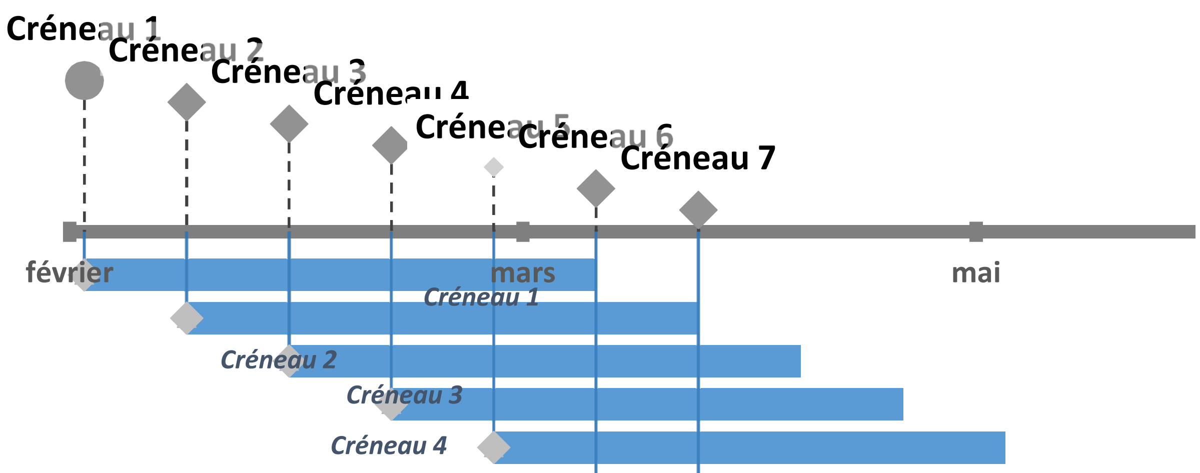
Exercice 2. Plan optimal pour évaluer des variétés de salades (inspiré d'une situation réelle en entreprise : SudExpé, par Alexandra Candeille)

Une entreprise d'expérimentation agricole souhaite réaliser des essais variétaux pour fournir des références technico-économiques aux agriculteurs. On souhaite évaluer les performances de 30 variétés de salades différentes, sous serre, avec 7 dates différentes de semis (créneaux). Comme les salades sont des plantes à croissance rapide, les 2 premiers créneaux peuvent être enchainés directement avec les 2 derniers créneaux, ce qui permet de rendre la place disponible 2 fois (cf. schéma ci-dessous). Pour tester une variété, il faut 7,5m². Or, l'entreprise ne peut pas rendre disponible plus de 1000m² de serre pour cette expérience.

Quel plan d'expérience pouvez-vous lui proposer ? Est-ce que ce plan pourra estimer l'interaction variété x date de semis ?

Idéalement, le plan complet devrait comporter $30 \times 7 = 210$ essais.

Comme la durée de croissance de la salade le permet, on peut enchainer le créneau 6 directement après la fin du créneau 1, et le créneau 7 après le créneau 2. Ainsi, en termes de place, on n'a besoin que de l'équivalent de 5 créneaux (cf. schéma ci-dessous)



Surface de serre remplie (m²)	*200*	*400*	*600*	*800*	*1000*	*1000*	*1000*	*800*	*600*	*400*	*200*

Pour avoir un plan orthogonal, on va tester chaque créneau (c'est-à-dire chaque date de semis) le même nombre de fois. C'est-à-dire qu'à chaque nouveau créneau, on va occuper 200m² (= $\frac{1000m^2}{5}$) (comme expliqué ci-dessus).

Pour estimer les coefficients du modèle sans interaction, il faut au minimum : $1 + (30 - 1) + (7 - 1) = 36$ essais.

L'objectif maintenant est de déterminer quelles variétés planter à chaque créneau. Puisqu'une variété occupe 7,5m² et qu'on utilise 200m² à chaque fois, on va pouvoir tester $200 \div 7{,}5 = 26$ variétés à chaque fois. Or, on en a 30 à tester en tout. Nous allons donc réaliser $26 \times 7 = 182$ essais.

Pour construire le plan, il suffit de randomiser les variétés dans les créneaux, c'est-à-dire sélectionner les 182 essais parmi 210 pour optimiser le plan le plus possible. C'est ce que nous allons faire ici avec la fonction `opt.federov`.

```
library(DoE.base)
plan<-fac.design(nlevels=c(30,7), factor.names=c("variete","creneau"))
planopt<-optFederov(~.,data=plan,nTrials=182,eval=TRUE)
```

```
View(planopt$design)

# je n'affiche pas le résultat ici pour des raisons de place, mais la commande ci-dessous permet de vérifier l'allure générale du plan
#on verifie l'orthogonalité
table(planopt$design[,c(1,2)])
       creneau
variete 1 2 3 4 5 6 7
     1  1 1 1 1 1 0 1
     2  0 1 1 1 1 1 1
     3  1 1 1 1 1 0 1
     4  1 1 0 1 1 1 1
     5  1 1 1 1 1 1 0
     6  1 1 1 0 1 1 1
     7  1 1 1 1 1 1 0
     8  1 0 1 1 1 1 1
     9  1 1 0 1 1 1 1
     10 1 1 1 0 1 1 1
     11 1 0 1 1 1 1 1
     12 1 1 1 1 1 0 1
     13 1 1 1 1 1 1 0
     14 1 1 1 1 1 1 1
     15 1 0 1 1 1 1 1
     16 1 1 1 1 1 1 1
     17 0 1 1 1 1 1 1
     18 1 1 1 0 1 1 1
     19 1 1 1 1 0 1 1
     20 1 1 1 0 1 1 1
     21 1 1 1 1 0 1 1
     22 1 1 1 1 0 1 1
     23 1 1 1 1 1 1 0
     24 0 1 1 1 1 1 1
     25 0 1 1 1 1 1 1
     26 1 1 0 1 1 1 1
     27 1 0 1 1 1 1 1
     28 1 1 0 1 1 1 1
     29 1 1 1 1 1 0 1
     30 1 1 1 1 0 1 1
```

Le plan n'est pas orthogonal, mais chaque variété est testée sur 6 des 7 créneaux et chaque créneau est testé pour 26 variétés sur 30. C'est donc un plan non-orthogonal mais équilibré.

Malheureusement, ce plan ne permet pas d'estimer l'interaction entre les variétés de salades et les dates de semis, facteur qu'il est pourtant intéressant d'étudier en agronomie. Pour pouvoir estimer l'interaction entre les variétés et les dates de semis, il faudrait faire $PPCM(30 \times 7) = PPCM(210) = 210$ essais, c'est-à-dire le plan complet.

Le plan complet comporte $30 \times 7 = 210$ essais. Cela signifie que pour le réaliser, il aurait fallu mobiliser $210 \times \frac{5}{7} \times 7{,}5 = 1125$m² de serres. Dans cette situation, la meilleure chose à faire serait de recommander à l'entreprise de mobiliser un peu plus de surface de serres pour cette expérience, ce qui lui permettrait de gagner en qualité d'estimation.

Exercice 3. Photographie sur les réseaux sociaux (exemple réel, par Léa Pautrel)

J'ai décidé de me lancer dans la photographie professionnelle il y a quelques mois. Afin d'avoir des clients, il est aujourd'hui nécessaire de poster du contenu sur les réseaux sociaux pour se faire connaître. Le but de cette expérience est de voir quelles sont les photos qui sont les plus populaires, lorsqu'elles sont postées sur mon compte Instagram. La popularité des photos sera évaluée par le nombre de cœurs sur la publication. Les critères qui seront évalués, adaptés aux photos qui sont faites habituellement, sont les suivants. Il y a 1 facteur à 2 modalités, 2 facteurs à 3 modalités, et 2 facteurs à 4 modalités.

Facteur	Code	Modalité 1	Modalité 2	Modalité 3	Modalité 4
Nombre d'images	N	Image unique	Diptyque		
Format	F	Carré	Vertical	Horizontal	
Couleurs dominantes	C	Froides (bleu, vert)	Chaudes (rouge, orange)	Noir et blanc	
Sujet	S	Portrait créatif	Portrait naturel	Nature	Autre
Heure de publication	H	Matin (6h → midi)	Après-midi (midi → 18h)	Soirée (18h → minuit)	Nuit (minuit→ 6h)

1. Déterminer le nombre de photos à poster pour :

a. Avoir le plan complet

Il suffit de multiplier le nombre de modalités par facteur entre eux :

$$2 \times 3 \times 3 \times 4 \times 4 = \mathbf{288}$$

Il faut poster 288 photos pour avoir le plan complet.

b. Avoir l'orthogonalité entre les facteurs 2 à 2 avec le moins de photos possible

Afin que les facteurs soient orthogonaux deux à deux, il faut :

- $N \perp F\,;\; N \perp C \rightarrow 2 \times 3 = \mathbf{6}$
- $N \perp S\,;\; N \perp H \rightarrow 2 \times 4 = \mathbf{8}$
- $F \perp S\,;\; F \perp H\,; C \perp S\,; C \perp H \rightarrow 3 \times 4 = \mathbf{12}$
- $F \perp C \rightarrow 3 \times 3 = \mathbf{9}$
- $S \perp H \rightarrow 4 \times 4 = \mathbf{16}$

On recherche donc le PPCM (plus petit commun multiple) de ces nombres.

$$PPCM\,(6, 8, 9, 12, 16) \;=\; \mathbf{144}$$

Il faut donc faire 144 essais au minimum, soit poster 144 photos, pour avoir un plan d'expérience parfaitement orthogonal afin d'estimer les effets principaux.

c. Estimer les effets principaux avec le moins de photos possible

Il s'agit du nombre minimum de paramètres à estimer :

- 1 pour la constante
- $(n-1)$ pour les facteurs à n modalités

Soit : $1 + (2-1) + 2 \times (3-1) + 2 \times (4-1) = \mathbf{12}$

Il faut faire poster au minimum 12 photos.

2. On s'intéresse d'abord uniquement aux variables n'étant pas en lien avec le format de l'image. On souhaite donc étudier uniquement l'effet des variables 'Couleurs dominantes' (*C, 3 modalités*) ; 'Sujet' (*S, 4 modalités*) et 'Heure de publication' (*H, 4 modalités*). On souhaite poster seulement 16 photos. Construire un plan d'expérience à la main.

On construit d'abord un carré latin de taille 4 à 3 facteurs (C, S, H) à 4 modalités chacun.

	S1	S2	S3	S4
C1	H1	H2	H3	H4
C2	H2	H3	H4	H1
C3	H3	H4	H1	H2
C4	H4	H1	H2	H3

→

PHOTO	C	S	H
1	1	1	1
2	1	2	2
3	1	3	3
4	1	4	4
5	2	1	2
6	2	2	3
7	2	3	4
8	2	4	1

9	3	1	3
10	3	2	4
11	3	3	1
12	3	4	2
13	4	1	4
14	4	2	1
15	4	3	2
16	4	4	3

La variable 'Couleurs dominantes' n'ayant que 3 modalités, on compresse deux niveaux. Les couleurs dominantes que j'utilise le plus étant les couleurs froides, j'ai choisi de compresser les variables 1 et 2 qui formeront la modalité 'Froides (bleu, vert)'.

Le tableau ci-dessus a été copié sur Excel et modifié de la manière suivante :

COLONNE 1	$= SI((B2 = 1) + OU(B2 = 2); "Froides"; SI(B2 = 3; "Chaudes"; "N\&B"))$
COLONNE 2	$= SI(C2 = 1; "Portrait\ créatif";\ SI(C2 = 2; "Portrait\ naturel"; SI(C2 = 3; "Nature"; "Autre")))$
COLONNE 3	$= SI(D2 = 1; "Matin";\ SI(D2 = 2; "Après - midi"; SI(D2 = 3; "Soirée"; "Nuit")))$

Le plan d'expérience final obtenu est le suivant :

PHOTO	C	S	H
1	Froides	Portrait créatif	Matin
2	Froides	Portrait naturel	Après-midi
3	Froides	Nature	Soirée
4	Froides	Autre	Nuit
5	Froides	Portrait créatif	Après-midi
6	Froides	Portrait naturel	Soirée
7	Froides	Nature	Nuit
8	Froides	Autre	Matin
9	Chaudes	Portrait créatif	Soirée
10	Chaudes	Portrait naturel	Nuit
11	Chaudes	Nature	Matin
12	Chaudes	Autre	Après-midi
13	N&B	Portrait créatif	Nuit
14	N&B	Portrait naturel	Matin
15	N&B	Nature	Après-midi
16	N&B	Autre	Soirée

3. **Afin d'avoir des résultats rapidement, on souhaite désormais étudier les 5 variables en postant seulement 16 photos. On cherche plan d'expérience optimal. Sur R construire le plan complet (fonction fac.design du package DoE.base) puis construire le plan optimal (fonction optFederov du package AlgDesign).**

```
# Construction du plan complet (fac.design)
library(DoE.base)
plan_complet <- fac.design(nlevels=c(2, 3, 3, 4, 4), nfactors=5)

# Plan optimal (optFederov)
# 12 essais : on ne peut pas estimer les effets quadratiques ni les interactions (pas assez de
ddl).
library(AlgDesign)
plan_opt = optFederov(~., # plan orthogonal
                      plan_complet, # à partir du plan décrit ci-dessus
                      nTrials=16, # en 16 essais
                      criterion="D")  # critère de D-optimalité
plan_opt


$D
[1] 0.3308785

$A
[1] 3.502603

$Ge
```

```
[1] 0.674

$Dea
[1] 0.617

$design
    A B C D E
38  2 1 2 1 3
52  1 1 3 2 1
54  1 2 2 3 4
67  1 3 2 2 3
118 2 1 3 4 4
121 2 2 1 1 1
139 2 3 2 4 1
143 2 3 1 3 3
179 1 3 1 1 2
180 1 3 3 1 4
198 1 1 1 4 2
205 2 3 3 3 2
242 1 1 2 3 1
252 2 1 1 2 4
267 1 2 3 4 3
277 2 2 2 2 2

$rows
 [1]  38  52  54  67 118 121 139 143 179 180 198 205 242 252 267 277
```

4. **Calculer la matrice $(X'X)^{-1}$ associé à ce plan d'expérience et commenter les résultats.**

```
options(contrasts = c("contr.sum", "contr.sum"))
# on considère que la somme des coefficients est égale à 0

X <- model.matrix(~., data = plan_opt$design)
round(solve(t(X)%*%X), digits = 2)
```

```
            (Intercept)    A1   B.L   B.Q   C.L   C.Q   D.L   D.Q   D.C   E.L   E.Q   E.C
(Intercept)        0.07  0.00  0.00 -0.03  0.00  0.01  0.00  0.00  0.00  0.00  0.00  0.00
A1                 0.00  0.06  0.00  0.00  0.02  0.00 -0.01  0.00  0.00  0.00  0.00  0.00
B.L                0.00  0.00  0.20  0.00 -0.02  0.00  0.04  0.00  0.07  0.00  0.08  0.00
B.Q               -0.03  0.00  0.00  0.23  0.00 -0.04  0.00  0.01  0.00  0.00  0.00  0.01
C.L                0.00  0.02 -0.02  0.00  0.23  0.00 -0.06  0.00 -0.03 -0.04 -0.05  0.02
C.Q                0.01  0.00  0.00 -0.04  0.00  0.21  0.00 -0.06  0.00 -0.03  0.00 -0.06
D.L                0.00 -0.01  0.04  0.00 -0.06  0.00  0.27  0.00  0.02  0.01  0.02  0.00
D.Q                0.00  0.00  0.00  0.01  0.00 -0.06  0.00  0.27  0.00  0.01  0.00  0.02
D.C                0.00  0.00  0.07  0.00 -0.03  0.00  0.02  0.00  0.27  0.00  0.03  0.00
E.L                0.00  0.00  0.00  0.00 -0.04 -0.03  0.01  0.01  0.00  0.26  0.01  0.01
E.Q                0.00  0.00  0.08  0.00 -0.05  0.00  0.02  0.00  0.03  0.01  0.29  0.00
E.C                0.00  0.00  0.00  0.01  0.02 -0.06  0.00  0.02  0.00  0.01  0.00  0.27
```

Il y a des valeurs différentes de 0 en dehors de la diagonale, l'orthogonalité n'est donc pas parfaite. Cependant, ces valeurs sont toutes inférieures à 0,1 et généralement bien inférieures aux valeurs présentes sur la diagonale. Il y a donc de nombreuses confusions entre certains facteurs, mais elles sont négligeables et le plan d'expérience reste intéressant.

Exercice 4. Temps de croissance d'un pied de tomate (exemple fictif, par Zoé Wante)

Nous cherchons ici à minimiser le temps de croissance d'un pied de tomate. On s'intéressera à l'effet de 4 facteurs : le lieu de pouce des racines, le taux de matière organique dans la terre, la température moyenne et enfin le taux d'humidité. Les racines peuvent pousser dans l'eau ou dans la terre, le taux de matière organique est de 1,5%, 2% ou 2,5%, la température moyenne est de 16, 19 ou 21ºC et enfin le taux d'humidité de 20%, 30%, 40%, 50% ou 60%.

1. **Combien faut-il d'essais pour créer un plan complet ?**

Nous sommes face à un plan à 4 facteurs avec 2, 3, 3 et 5 modalités. Le plan complet est donc de : 2 x 3^2 x 5 = 90 essais. On remarquera qu'il s'agit également du plus petit commun multiple et donc que ce plan total est le plus petit permettant l'orthogonalité entre les facteurs 2 à 2.

2. **Citer au moins un critère que l'on utilise pour s'approcher d'un plan optimal.**

Nous chercherons à minimiser le déterminant de la matrice $({}^tXX')^{-1}$ (D-optimalité) ou minimiser la variance des coefficients du modèle (A-optimalité). Le critère de D-optimalité est le plus utilisé.

3. **Construire, sans afficher, un plan complet à l'aide de R.**

Voici le code à utiliser :

```
plan.complet <- fac.design(nlevels=c(2,3,3,5), factor.names= c("lieu_de_pouce ","
taux_de_MO","température","humidité")
```

4. **Ecrire une fonction Plan qui prend en argument un plan complet, un nombre d'essais minimum, un nombre d'essais maximum et un déterminant à approcher. Cette fonction renvoie le nombre d'essais nécessaires pour former un plan optimal tel que le déterminant de la matrice $({}^tXX)^{-1}$ soit inférieur à celui annoncé et avec le moins d'essais possibles sachant que le nombre d'essais doit être compris entre le minimum et le maximum désirés. Cette fonction retourne également la matrice $({}^tXX)^{-1}$ et son déterminant. Testez cette fonction sur le plan créé à la question 3 et donnez le nombre d'essais minimum pour que le déterminant soit le plus proche de 10^{-7}**

```
Plan <- function(plan.complet, min, max, det){
  detmin <- 10^5
  essais = 0
  for (i in min:max) {
    plan.opt<-optFederov(~.,plan.complet,nTrials=i,criterion ="D") #Je crée le plan optimal pour chaque i avec le critere de D-optimalité
    Xopti=model.matrix(~ . , plan.opt$design)
    XX <- solve(t(Xopti)%*%Xopti)
    determinant <- det(XX)  #On cherche a le minimiser
    if (determinant < detmin){
        if (determinant >  det){
          mini <- determinant
          essais <- i}}
 }
  print(round(XX, 2)) #On affiche la matrice
  resultats = c(essais, mini)
  return(resultats)}
 Plan(plan.complet, 10, 90, 10^(-7))
```

Après l'avoir lancé la fonction nous trouvons qu'il faut 15 essais pour approcher $\det(({}^tXX)^{-1}) = 10^{-7}$.

Exercice 5. Comment organiser la meilleure animation de l'année ? (thématique réelle, par M. Madrolle)

La thématique de cet exercice est réelle, mais les expériences n'ont pas été effectuées.

Une personne organise une animation dans un centre commerciale. Le but de cette action est de récolter des fonds pour une association et d'attirer le plus de monde possible. Selon elle, l'attractivité de son stand peut dépendre de plusieurs facteurs à plusieurs modalités :

- Couleur dominante : « Jaune », « Rouge », « Bleu »
- Personnels employé ce jour : «Team 1», «Team 2», «Team 3»
- Emplacement : « Proche d'une entrée » ou « Au centre de la galerie »
- Type d'animation : « Tombola », « Débat », « Stand photo », « Dégustation »

L'objectif de cette association est de proposer les meilleures animations le plus vite possible, afin de récolter des fonds efficacement.

1. **Quel est le nombre d'essais associés au plan complet ? Combien faut-il d'essais au minimum si on veut orthogonalité entre les facteurs 2 à 2 ?**

Ici on s'intéresse à l'effet de 4 facteurs (couleur, employés, emplacement, animation) sur la réponse Y (attractivité du stand). Les facteurs A et B ont 3 modalités, C a 2 modalités et D a 4 modalités.

Le nombre d'essais associé au plan complet est : 2^1*3^2*4^1=2*9*4=72 essais

Nombre d'essais au minimum pour avoir l'orthogonalité entre les facteurs 2 à 2 :

A et B orthogonaux à C : 2*3=6 essais

A et B orthogonaux à D : 3*4=12 essais
C orthogonal à D : 2*4=8 essais
A orthogonal à B : 3*3=9 essais
N>=PPC(6,12,8,9)=72

Il en faut au minimum 72 essais pour avoir orthogonalité des facteurs 2 à 2.

2. **Construire un plan complet, puis un plan optimal en 9 essais.**

Construction du plan d'expérience complet sur R :

```
library(DoE.base)
PlanComplet <-fac.design(nlevels=c(3,3,2,4),factor.names=(c("couleur", "employés", "emplacement",
"animation"))) #plan complet
PlanComplet

class=design, type= full factorial

   couleur employés emplacement animation
1        2        1           1         1
2        1        2           2         1
3        3        2           2         1
4        3        2           1         2
5        2        2           2         1
6        2        1           1         3
7        1        3           1         4
8        2        3           2         4
9        3        1           1         4
[ . . .]

66       1        3           2         1
67       1        3           1         1
68       3        1           1         3
69       2        2           1         3
70       1        1           2         1
71       1        3           1         3
72       1        3           2         3
class=design, type= full factorial

#Plan otimal en 9 essais
library(AlgDesign) #package
set.seed(1234)   # permet de retrouver les mêmes résultats
planOpt<-optFederov(~.,data=PlanComplet,nTrials=9)
planOpt

$D
[1] 0.3145064

$A
[1] 4.962963

$Ge
[1] 0.2

$Dea
[1] 0.018

$design
   couleur employés emplacement animation
3        2        1           2         3
17       2        3           1         2
32       2        2           2         4
35       3        2           1         3
38       1        2           2         1
41       1        3           2         3
45       3        3           2         4
53       1        1           1         4
64       1        2           2         2

$rows
[1]  3 17 32 35 38 41 45 53 64
```

On peut notifier que ce plan d'expérience est un plan asymétrique qui peut se construire à la main. C'est un plan à 2,3 et 4 niveaux.

3. **Avec 9 essais, quel modèle peut-on considérer ? Préciser les degrés de liberté associés à chaque facteur.**

Avec 9 essais nous pouvons analyser le model d'analyse de variance à 4 facteurs.

Nombre de paramètres à estimer : 1+2*(3-1) +1*(2-1) +1*(4-1) =9 ddl en 9 essais.

On va faire 9 essais, donc on n'a plus de ddl pour estimer la résiduelle = 0 ddl

On ne peut donc pas la calculer, le modèle est complètement saturé.

4. **Calculer la matrice de dispersion associée à ce modèle. Commentez la propriété de ce plan.**

```
#Matrice de dispersion à 9 essais :
library(AlgDesign) #package
set.seed(12345)
planOpt<-optFederov(~couleur+employés+emplacement+animation,data=PlanComplet,nTrials=9)        #9
essais
XOpti<-model.matrix(~couleur+employés+emplacement+animation,planOpt$design)      #les      effets
quadratiques de la variable design du plan optimal
round(solve((t(XOpti)%*%XOpti)),2)
```

```
           (Intercept) couleur.L couleur.Q employés.L employés.Q emplacement1 animation.L animation.Q animation.C
(Intercept)   0.13     -0.09      0.02      0.01      -0.03       0.01       -0.01        0.05        0.03
couleur.L    -0.09      0.50     -0.14      0.00       0.14       0.00        0.00       -0.18        0.00
couleur.Q     0.02     -0.14      0.48     -0.03       0.06      -0.06        0.06        0.28       -0.13
employés.L    0.01      0.00     -0.03      0.44      -0.03       0.08       -0.21       -0.04       -0.07
employés.Q   -0.03      0.14      0.06     -0.03       0.48      -0.06        0.06        0.18       -0.13
emplacement1 0.01      0.00     -0.06      0.08      -0.06       0.14       -0.04       -0.07       -0.01
animation.L  -0.01      0.00      0.06     -0.21       0.06      -0.04        0.60        0.07        0.03
animation.Q   0.05     -0.18      0.28     -0.04       0.18      -0.07        0.07        0.72       -0.16
animation.C   0.03      0.00     -0.13     -0.07      -0.13      -0.01        0.03       -0.16        0.51
```

En 9 essais, l'orthogonalité n'est pas parfaite. On remarque qu'il y a des confusions de quasiment tous les facteurs et la constante.

5. **Faites la même chose avec un plan d'expérience optimal et la matrice de dispersion pour le plan à 72 essais. Commentez le résultat.**

```
PlanComplet <-fac.design(nlevels=c(3,3,2,4),factor.names=(c("couleur", "employés",
"emplacement", "animation"))) #plan complet
library(AlgDesign) #package
set.seed(124)
planOpt<-optFederov(~couleur+employés+emplacement+animation,data=PlanComplet,nTrials=72) #72
essais
XOpti<-model.matrix(~couleur+employés+emplacement+animation,planOpt$design) #les effets
quadratiques de la variable design du plan optimal
round(solve((t(XOpti)%*%XOpti)),2)
```

```
             (Intercept) couleur.L couleur.Q employés.L employés.Q emplacement1 animation.L animation.Q animation.C
(Intercept)        0.01      0.00      0.00       0.00       0.00         0.00        0.00        0.00        0.00
couleur.L          0.00      0.04      0.00       0.00       0.00         0.00        0.00        0.00        0.00
couleur.Q          0.00      0.00      0.04       0.00       0.00         0.00        0.00        0.00        0.00
employés.L         0.00      0.00      0.00       0.04       0.00         0.00        0.00        0.00        0.00
employés.Q         0.00      0.00      0.00       0.00       0.04         0.00        0.00        0.00        0.00
emplacement1       0.00      0.00      0.00       0.00       0.00         0.01        0.00        0.00        0.00
animation.L        0.00      0.00      0.00       0.00       0.00         0.00        0.06        0.00        0.00
animation.Q        0.00      0.00      0.00       0.00       0.00         0.00        0.00        0.06        0.00
animation.C        0.00      0.00      0.00       0.00       0.00         0.00        0.00        0.00        0.06
```

En 72 essais, le plan est bien orthogonal, il est donc de bonne qualité. On peut remarquer que le plan optimal correspond au plan au plan complet. Il faut donc faire tous les essais pour avoir la meilleure interprétation.

Exercice 6. Dispositif expérimental en horticulture (exemple fictif, par Mélanie Demeure)

On cherche à évaluer le comportement de 6 variétés de Chrysanthèmes (notées V1 à V6) en pot, dans une serre. On aimerait tester 5 doses d'azotes (noté N1 à N5). Les semis peuvent avoir été effectuer à 2 dates différentes (D1 et D2). Enfin, on estime que la situation dans la serre n'est pas homogène. On la divise en 4 blocs, dont les variables de température et d'ensoleillement seront plus homogène (B1 à B4).

1. **Quel est le nombre d'essais associés au plan complet ? Combien faut-il d'essais au minimum si on veut l'orthogonalité entre les facteurs 2 à 2 ?**

On a 4 facteurs avec respectivement 6, 5, 4 et 2 modalités. Le plan complet contient 6*5*4*2=240 essais. Le PPCM (2,4,5,6) = 60. On devrait donc pouvoir trouver un plan orthogonal à 60 essais.

2. **Après avoir construit un plan complet (fonction fac.design du package DoE.base), construire un plan orthogonal avec le nombre d'essais déterminés en question 1 (fonction optFederov du package AlgDesign)**

On utilise les lignes de codes suivantes :

```
library(AlgDesign)
library(DoE.base)
complet<-fac.design(nlevels = c(2,4,5,6),factor.names = c("Date","Bloc","Azote","Var"))
orthogonal<-optFederov(~.,data=complet,nTrials=60)
```

3. **Calculer la matrice de dispersion pour le plan. Cela correspond-t-il à ce que vous attendiez ?**

La matrice de dispersion est obtenue ci-après.

```
options(contrasts = c("contr.sum", "contr.sum"))
X <- model.matrix(~.,data=orthogonal$design)
round(solve(t(X)%*%X),2)
            (Intercept) Date1 Bloc.L Bloc.Q Bloc.C Azote.L Azote.Q Azote.C Azote^4 Var.L Var.Q Var.C Var^4 Var^5
(Intercept)        0.02  0.00   0.00   0.00   0.00    0.00    0.00    0.00    0.00  0.00  0.00   0.0  0.00  0.00
Date1              0.00  0.02   0.00   0.00   0.00    0.00    0.00    0.00    0.00  0.00  0.00   0.0  0.00  0.00
Bloc.L             0.00  0.00   0.07   0.00   0.00    0.00    0.00    0.00    0.00  0.00  0.01   0.0  0.00  0.00
Bloc.Q             0.00  0.00   0.00   0.07   0.00    0.00    0.00    0.00    0.00  0.00  0.00   0.0  0.01 -0.01
Bloc.C             0.00  0.00   0.00   0.00   0.07    0.00    0.00    0.00    0.00 -0.01  0.00   0.0  0.01  0.00
Azote.L            0.00  0.00   0.00   0.00   0.00    0.08    0.00    0.00    0.00  0.00  0.00   0.0  0.00  0.00
Azote.Q            0.00  0.00   0.00   0.00   0.00    0.00    0.08    0.00    0.00  0.00  0.00   0.0  0.00  0.00
Azote.C            0.00  0.00   0.00   0.00   0.00    0.00    0.00    0.08    0.00  0.00  0.00   0.0  0.00  0.00
Azote^4            0.00  0.00   0.00   0.00   0.00    0.00    0.00    0.00    0.08  0.00  0.00   0.0  0.00  0.00
Var.L              0.00  0.00   0.00   0.00  -0.01    0.00    0.00    0.00    0.00  0.10  0.00   0.0  0.00  0.00
Var.Q              0.00  0.00   0.01   0.00   0.00    0.00    0.00    0.00    0.00  0.00  0.10   0.0  0.00  0.00
Var.C              0.00  0.00   0.00   0.00   0.00    0.00    0.00    0.00    0.00  0.00  0.00   0.1  0.00  0.00
Var^4              0.00  0.00   0.00   0.01   0.01    0.00    0.00    0.00    0.00  0.00  0.00   0.0  0.10  0.00
Var^5              0.00  0.00   0.00  -0.01   0.00    0.00    0.00    0.00    0.00  0.00  0.00   0.0  0.00  0.10
```

En effet, les effets principaux semblent quasiment orthogonaux (les valeurs hors diagonales sont tellement petites – de l'ordre de 10^{-19} – que l'on peut les considérer comme nulles). On pourrait estimer les effets principaux sans confusions.

4. **Après évaluation des coûts de l'expérience, on préfère se limiter à 16 essais. Après avoir construit un plan complet (fonction fac.design du package DoE.base), construire un plan optimal en 16 essais (fonction optFederov du package AlgDesign).**

On utilise les lignes suivantes pour construire le plan. On utilise l'algorithme de Federov à partir du plan complet. (A noter que le plan pourra être différent d'une itération du programme à l'autre).

```
library(AlgDesign)
library(DoE.base)
set.seed(123)
complet<-fac.design(nlevels = c(2,4,5,6),factor.names = c("Date","Bloc","Azote","Var"))
optimal<-optFederov(~.,data=complet,nTrials=16)
optimal$design

    Date Bloc Azote Var
9      1    2     1   2
33     1    3     5   3
74     2    1     5   6
92     2    3     3   4
107    1    4     4   4
132    2    2     5   1
145    2    4     2   2
148    1    1     1   1
151    2    1     4   5
163    2    4     1   5
173    2    1     2   4
176    2    3     4   6
183    2    2     3   3
192    1    1     3   2
212    1    4     3   6
233    1    2     2   5
```

5. **Avec 16 essais, quel modèle d'analyse peut-on considérer ? Donner les degrés de libertés associés à chaque facteur.**

On pourra alors faire une analyse de variance sur les effets principaux. On aura alors à estimer 5 paramètres : μ et les 4 variables. Les degrés de liberté associés sont 1 pour μ, 1 pour la variable Date, 3 pour la variable Bloc, 4 pour la variable Azote et 5 pour la variable Variété. Cela fait donc un total de 14 degrés de liberté pour l'estimation des paramètres. Il reste 2 ddl pour la résiduelle.

6. **Calculer la matrice des effets associée à votre modèle et votre plan, puis calculer la matrice de dispersion. Commenter les propriétés de ce plan.**

De la même manière que précédemment, on utilise les codes suivants pour afficher les matrices des effets et de dispersion.

```r
options(contrasts = c("contr.sum", "contr.sum"))
X <- model.matrix(~.,data=optimal$design)
round(solve(t(X)%*%X),2)
```

	(Intercept)	Date1	Bloc.L	Bloc.Q	Bloc.C	Azote.L	Azote.Q	Azote.C	Azote^4	Var.L	Var.Q	Var.C	Var^4	Var^5
(Intercept)	0.07	-0.01	0.02	-0.01	-0.02	-0.01	-0.01	0.00	-0.03	-0.03	0.02	0.00	0.02	-0.03
Date1	-0.01	0.07	0.00	0.01	0.01	-0.02	0.01	-0.02	0.01	-0.03	-0.01	0.03	0.00	0.00
Bloc.L	0.02	0.00	0.28	0.01	-0.07	0.03	-0.01	0.05	0.02	-0.11	0.07	-0.07	-0.05	-0.03
Bloc.Q	-0.01	0.01	0.01	0.33	0.05	0.05	0.01	0.00	0.01	-0.05	-0.07	0.04	0.09	-0.12
Bloc.C	-0.02	0.01	-0.07	0.05	0.40	0.07	0.04	-0.08	0.00	0.11	-0.08	0.07	0.12	0.07
Azote.L	-0.01	-0.02	0.03	0.05	0.07	0.53	0.02	0.00	0.08	-0.06	-0.06	-0.11	-0.24	0.13
Azote.Q	-0.01	0.01	-0.01	0.01	0.04	0.02	0.46	-0.01	0.09	0.20	-0.20	0.06	-0.02	0.16
Azote.C	0.00	-0.02	0.05	0.00	-0.08	0.00	-0.01	0.38	0.03	0.03	0.07	-0.08	-0.05	0.00
Azote^4	-0.03	0.01	0.02	0.01	0.00	0.08	0.09	0.03	0.39	0.09	-0.04	-0.14	-0.15	0.11
Var.L	-0.03	-0.03	-0.11	-0.05	0.11	-0.06	0.20	0.03	0.09	0.60	-0.14	0.05	0.01	0.14
Var.Q	0.02	-0.01	0.07	-0.07	-0.08	-0.06	-0.20	0.07	-0.04	-0.14	0.54	-0.11	-0.02	-0.02
Var.C	0.00	0.03	-0.07	0.04	0.07	-0.11	0.06	-0.08	-0.14	0.05	-0.11	0.49	0.13	-0.06
Var^4	0.02	0.00	-0.05	0.09	0.12	-0.24	-0.02	-0.05	-0.15	0.01	-0.02	0.13	0.58	-0.14
Var^5	-0.03	0.00	-0.03	-0.12	0.07	0.13	0.16	0.00	0.11	0.14	-0.02	-0.06	-0.14	0.56

On se rend compte qu'effectivement, le plan à 16 essais comprend de nombreuses confusions entre les effets, puisque la matrice n'est pas diagonale, et il faudra le prendre en compte dans l'analyse et l'interprétation des résultats.

Exercice 7. Recherche des meilleures conditions pour faire pousser une plante (exemple réel, par Soline Mounsif)

L'entreprise X, vient de recevoir de nouvelles graines.
Elle aimerait déterminer les meilleures conditions de pousse pour optimiser sa culture, ses rendements.
La commercialisation de ces plantes n'est prévue que pour l'année prochaine. Elle a donc le temps de réaliser quelques tests.
Elle décide de s'intéresser à 5 facteurs :

- Facteur A : Taille de la plante (oui ou non)
- Facteur B : Palissage (oui ou non)
- Facteur C : Volume du pot (5L ou 20L)
- Facteur D : Engrais (billes NPK, fumier, compost, sang)
- Facteur E : Substrat (terreau, terreau+perlite, terreau+vermiculite, sol, sable)

1. **Quel est le nombre d'essais associé au plan complet ?**

2^3 x 4 x 5 = 160 essais sont nécessaires

2. **Combien d'essais faudrait-il au minimum pour respecter l'orthogonalité des facteurs 2 à 2 ?**

Pour estimer sans confusions d'effets, il faut respecter l'orthogonalité des facteurs 2 à 2 :
A ⊥ B, A ⊥ C, B ⊥ C donc minimum 2 x 2 = 4 essais
A ⊥ D, B ⊥ D, C ⊥ D donc minimum 2 x 4 = 8 essais
A ⊥ E, B ⊥ E, C ⊥ E donc minimum 2 x 5 = 10 essais
D ⊥ E donc minimum 4 x 5 = 20 essais
Le nombre d'essais doit être supérieur ou égal au PPCM(4, 8, 10, 20), soit à 44.

3. **Les plannings des employés de l'entreprise étant déjà très chargés, la responsable de production décide arbitrairement de ne réaliser que 13 essais. Construisez un plan complet à l'aide de la fonction fac.design du package DoE.base, puis un plan optimal en 13 essais en utilisant la fonction optFederov du package AlgDesign.**

```r
library(DoE.base)
library(AlgDesign)
comp<-fac.design(nlevels=c(2,2,2,4,5))
res<-optFederov(~.,data=comp,nTrials=13,eval=TRUE)
res
```

```
$design
    A B C D E
9   2 1 2 2 1
14  2 1 1 2 3
25  2 2 2 4 4
30  2 1 1 3 5
39  2 1 1 4 2
44  2 1 2 1 5
59  1 1 1 1 4
86  1 1 2 3 2
99  2 2 1 1 2
100 1 2 1 2 5
108 1 1 1 4 1
148 2 2 1 3 1
153 1 2 2 1 3
```

4. **Quel modèle peut-on considérer ? Quels sont les degrés de libertés associés à celui-ci ?**

On réalise une analyse de variance à 5 facteurs.
Pour la constante : 1 ddl
Pour les 3 facteurs à 2 modalités : 3 x (2-1) = 3 ddl
Pour le facteur à 4 modalités : 1 x (4-1) = 4 ddl
Pour le facteur à 5 modalités : 1 x (5-1) = 5 ddl
Soit 13 ddl nécessaires au total pour estimer les facteurs uniquement.
Le plan est saturé. Il n'y a plus de ddl pour estimer les interactions et la résiduelle.

5. **Calculez la matrice des effets associée à ce modèle et à votre plan, puis calculez la matrice de dispersion. Commentez l'orthogonalité. Que dîtes-vous à la responsable ?**

```
options(contrasts = c("contr.sum", "contr.sum"))
X <- model.matrix(~ . , res$design)
inv<-t(X)%*%X
round(inv,2)
```

	(Intercept)	A1	B1	C1	D.L	D.Q	D.C	E.L	E.Q	E.C	E^4
(Intercept)	13.00	3.00	3.00	3.00	-0.22	-0.50	0.67	-0.32	-0.27	-0.95	0.36
A1	3.00	13.00	1.00	1.00	0.22	0.50	-0.67	-0.32	-0.27	-0.95	0.36
B1	3.00	1.00	13.00	1.00	0.22	0.50	-0.67	-0.32	-0.27	-0.95	0.36
C1	3.00	1.00	1.00	13.00	0.22	0.50	-0.67	-0.32	-0.27	-0.95	0.36
D.L	-0.22	0.22	0.22	0.22	3.05	0.11	-0.15	-0.07	0.66	-0.92	0.24
D.Q	-0.50	0.50	0.50	0.50	0.11	3.25	-0.34	-0.16	-1.20	-0.47	-0.06
D.C	0.67	-0.67	-0.67	-0.67	-0.15	-0.34	3.45	-0.49	-0.18	0.64	0.61
E.L	-0.32	-0.32	-0.32	-0.32	-0.07	-0.16	-0.49	2.50	-0.42	0.00	-0.23
E.Q	-0.27	-0.27	-0.27	-0.27	0.66	-1.20	-0.18	-0.42	2.64	0.00	-0.19
E.C	-0.95	-0.95	-0.95	-0.95	-0.92	-0.47	0.64	0.00	0.00	2.50	0.26
E^4	0.36	0.36	0.36	0.36	0.24	-0.06	0.61	-0.23	-0.19	0.26	2.76

Il y a des confusions dans cette matrice. Les covariances ne sont pas nulles. Il n'y a pas indépendance entre les facteurs
La responsable doit réaliser plus d'essais si elle veut que son étude soit pertinente.
Cela ne sert à rien de déployer de l'énergie pour seulement 13 essais qui ne seront pas concluants. Nous avons vu en question 2 que 44 essais minimum étaient nécessaires. Il faut lui conseiller de réaliser 44 essais ou plus.

6. **Refaire l'analyse en tenant compte de vos conseils, et vérifier la qualité de votre plan.**

```
library(DoE.base)
library(AlgDesign)
comp<-fac.design(nlevels=c(2,2,2,4,5))
res<-optFederov(~.,data=comp,nTrials=44,eval=TRUE)
res
```

```
$design
    A B C D E
4   2 2 2 3 2
8   1 1 1 4 2
9   1 2 1 1 1
13  1 1 2 2 2
17  2 1 2 2 5
[ . . . ]
148 1 2 1 1 5
149 1 2 2 2 3
156 2 1 1 2 4
158 1 1 2 2 1
```

```
options(contrasts = c("contr.sum", "contr.sum"))
X <- model.matrix(~ . , res$design)
inv<-t(X)%*%X
round(inv,2)
            (Intercept) A1 B1 C1   D.L   D.Q   D.C   E.L   E.Q   E.C   E^4
(Intercept)       44.00  0  0  0 -0.89  0.00 -1.79 -0.63 -1.60  1.26  0.48
A1                 0.00 44  0  0  0.00  0.00  0.00  0.00  0.00  0.00  0.00
B1                 0.00  0 44  0  0.00  0.00  0.00  0.00  0.00  0.00  0.00
C1                 0.00  0  0 44  0.00  0.00  0.00  0.00  0.00  0.00  0.00
D.L               -0.89  0  0  0 11.00 -0.89  0.00  0.14  0.36 -0.28 -0.11
D.Q                0.00  0  0  0 -0.89 11.00  0.45  0.00  0.00  0.00  0.00
D.C               -1.79  0  0  0  0.00  0.45 11.00  0.28  0.72 -0.57 -0.21
E.L               -0.63  0  0  0  0.14  0.00  0.28  8.20  0.17 -0.40  0.30
E.Q               -1.60  0  0  0  0.36  0.00  0.72  0.17  8.71 -0.34 -0.51
E.C                1.26  0  0  0 -0.28  0.00 -0.57 -0.40 -0.34  8.80 -0.60
E^4                0.48  0  0  0 -0.11  0.00 -0.21  0.30 -0.51 -0.60  9.49
```

La matrice obtenue avec 44 essais est presque orthogonale. Avec plus de 44 essais, l'orthogonalité sera respectée. Il n'y aura plus de confusion entre les facteurs. Ils seront indépendants.

Exercice 8. Plants de tomates dans une serre (contexte réel, par C. Fouilland)

Voici l'énoncé de mon exercice issu d'un exemple fictif. On s'intéresse à l'effet des conditions de culture sur les tomates cultivées.

La culture sous serre est beaucoup utilisée en agronomie puisqu'elle permet de mieux contrôler les conditions de culture. En recherche notamment, la culture sous serre permet d'étudier différents paramètres et leurs impacts sur les cultures. Dans cet exemple, des plants de tomates sont plantés dans une serre et on étudie plusieurs facteurs :

- F1 : Taux d'humidité dans la serre : sec (<30%), normal ou humide (>80%)
- F2 : Nombre de plantes par pot : 1, 5 ou 10
- F3 : Variété des graines plantées : Fandango, Noire de Crimée, Sungold ou Supersteak
- F4 : Température dans la serre : 5°C, 8°C, 11°C, 14°C, 17°C, 20°C ou 23°C.

1. **Calculer le nombre d'essais maximum et minimum (nécessaires si on veut l'orthogonalité entre les facteurs 2 à 2) pour construire un plan avec ces facteurs.**

Le nombre d'essais maximum correspond aux essais associés au plan complet. On obtient : 3^2 x 4 x 7 = 252 essais.

Le nombre d'essais minimum correspond au nombre d'essais minimum nécessaires si on veut l'orthogonalité entre les facteurs 2 à 2. On obtient : 1 + 2 x (3-1)+(4-1)+(7-1)= 14 essais.

2. **Existe-il un plan orthogonal avec un nombre d'essais inférieur au nombre d'essai maximal calculé dans la question précédente ? Si oui, donner le nom de ce plan.**

On veut l'orthogonalité des facteurs 2 à 2 donc il faut que toutes les combinaisons soient testées un même nombre de fois.

- **F1/F2 : donc minimum 3x3= 6 essais**
- **F1/F3 et F2/F3 : donc minimum 3 x 4 = 12 essais**
- **F1/F4 et F2/F4 : donc minimum 3x7 = 21 essais**
- **F3/F4 : donc minimum : 4 x 7 = 28 essais**

On calcule alors le PPCM de ces nombres : PPCM (6,12, 21, 28) = 84 essais. Donc il existe un plan orthogonal avec moins de 252 essais pour estimer les facteurs sans confusion d'effets. Ce plan est le plan orthogonal $L_{84}3^2 4^1 7^1$.

3. **La serre mise à disposition pour cette expérimentation ne permet pas de faire plus de 35 essais. Avec R, construire le plan complet puis le plan optimisé à 35 essais.**

On va réduire drastiquement le nombre d'essais par rapport au plan complet (35 au lieu de 252). On va donc détériorer la qualité du plan. Cependant, on cherche à déterminer les essais qui nous permettront néanmoins d'obtenir un plan de bonne qualité, c'est-à-dire « optimisé ». On utilise la fonction « `optFederov` » du package « `AlgDesign` » qui va donner le nom des lignes (essais) du plan complet à utiliser.

```
library(DoE.base)
PlanComplet<- fac.design(nlevels=c(3,3,4,7), factor.names= c("A","B","C","D"))
X <- model.matrix(~ . , PlanComplet)
qualiteX <- solve(t(X)%*%X)
determinant <- det(VARCOV)
library(AlgDesign)
PlanOptimisé<-optFederov(~.,Design.1,nTrials=35,criterion="D")
X2=model.matrix(~ . , PlanOptimisé$design)
```

4. **Comment peut-on vérifier la qualité de ce plan avec le logiciel R ? A partir de la sortie R donnée, peut-on dire que le plan est de bonne qualité ? Expliquer pourquoi.**

Pour vérifier la qualité d'un plan on calcule l'inverse du produit entre la matrice des essais et sa transposée : la matrice $(X'X)^{-1}$.

```
        det(t(X2)%*%X2)
        qualitéX2 <- solve(t(X2)%*%X2)
        round(qualitéX2, 2)
           (Intercept)   A.L   A.Q  B.L   B.Q   C.L   C.Q   C.C   D.L   D.Q   D.C  D^4   D^5   D^6
(Intercept)       0.03  0.00  0.00 0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00 0.00  0.00  0.00
A.L               0.00  0.09  0.00 0.00  0.00  0.00  0.00  0.00 -0.01  0.01 -0.01 0.01  0.00 -0.02
A.Q               0.00  0.00  0.09 0.00  0.00  0.00  0.01  0.00 -0.01 -0.01 -0.01 0.02  0.00  0.01
B.L               0.00  0.00  0.00 0.09  0.00  0.00  0.00  0.00  0.01  0.00  0.00 0.01  0.02  0.00
B.Q               0.00  0.00  0.00 0.00  0.09  0.00  0.00  0.00  0.01  0.01 -0.01 0.00 -0.01  0.02
C.L               0.00  0.00  0.00 0.00  0.00  0.12  0.00  0.00  0.02  0.01  0.01 0.00 -0.02  0.02
C.Q               0.00  0.00  0.01 0.00  0.00  0.00  0.12  0.00  0.00  0.00 -0.03 0.00  0.00  0.01
C.C               0.00  0.00  0.00 0.00  0.00  0.00  0.00  0.12 -0.01  0.01  0.00 0.02  0.01  0.02
D.L               0.00 -0.01 -0.01 0.01  0.01  0.02  0.00 -0.01  0.21  0.00  0.00 0.00  0.00  0.00
D.Q               0.00  0.01 -0.01 0.00  0.01  0.01  0.00  0.01  0.00  0.21  0.00 0.00  0.00  0.00
D.C               0.00 -0.01 -0.01 0.00 -0.01  0.01 -0.03  0.00  0.00  0.00  0.21 0.00  0.00  0.00
D^4               0.00  0.01  0.02 0.01  0.00  0.00  0.00  0.02  0.00  0.00  0.00 0.21  0.00  0.00
D^5               0.00  0.00  0.00 0.02 -0.01 -0.02  0.00  0.01  0.00  0.00  0.00 0.00  0.21  0.00
D^6               0.00 -0.02  0.01 0.00  0.02  0.02  0.01  0.02  0.00  0.00  0.00 0.00  0.00  0.22
```

Sur la sortie R, on peut voir que la matrice obtenue est presque diagonale, on note quelques valeurs non nulles en dehors de la diagonale mais ces valeurs restent très faibles. Les variances la matrice $(X'X)^{-1}$ sont faibles. Les covariances entre facteurs sont proches de 0, c'est-à-dire que même s'il n'y a pas toujours indépendance entre l'estimation des facteurs, la valeur de cette dépendance reste très faible. On peut donc dire que le plan est de bonne qualité.

Exercice 9. Le poisson fumé c'est bon (contexte réel, par Nicolas Andrialovanirina)

Après la combustion du bois, on a plus 200 composés qui sont produits. Ce sont ces composés qui donnent de l'arôme au poisson fumé. Nous nous intéressant ici aux quantités de phénol en mg/100g (donne des arômes, de la couleur au poisson, a également des actions antioxydantes et bactériostatiques) déposées dans la chair de poisson après un fumage. Cette quantité dépend des différents facteurs listés ci-dessous.

- température du foyer (tfo) : (1) 250°C, (2) 400°C.
- vitesse de l'air (va) : (1) 60m/minutes, (2) 120m/minutes.
- humidité du bois (hb) : (1) 5%, (2) 10%.
- nature du bois (nb) : (1) acacia, (2) oranger, (3) noix de coco.
- taille des particules du bois (tpb) : (1) 0,35mm, (2) 0,9mm, (3) 1,8mm.
- température fumée (tfu): (1) 60°C, (2) 71°C, (3) 82°C, (4) 93°C.
- humidité relative (Hr) de l'air : (1) 22%, (2) 40%, (3) 60%, (4) 80%.

Avec ces différents facteurs, on décide de réaliser des expériences (en plans optimaux) pour voir quelques modalités de ces facteurs produits de bon poisson fumé (c'est-à-dire avec le plus de phénol).

1. **Donnez le nombre d'essais sur le plan complet.**

$2^3 \times 3^2 \times 4^2 = 1152$ essais

2. **Quel est le nombre d'essais minimal que l'on peut faire ? Combien d'essais faut-il réaliser pour un plan orthogonal ?**

1+3(2-1)+2(3-1)+2(4-1) = 14 essais minimum

Pour un plan orthogonal : PPCM (4,6,8,9,12,16) = 144 essais

3. **Pour ne pas faire trop d'essais, on décide de faire le nombre d'essais minimum, pour un plan optimal. Réaliser ce plan avec R puis vérifier sa qualité.**

```
library(DoE.base)

# définition des essais potentiels possible (donc le plan complet)
plan.ex2 <- fac.design(nlevels = c(2,2,2,3,3,4,4), factor.names = c("tfo", "va", "hb", "nb",
"tpb", "tfu", "Hr"))
```

```
library(AlgDesign)
# construction du plan optimal avec les 14 essais minimaux
plan.opt <- optFederov(~., plan.ex2, nTrials = 14,  criterion = "D")
plan.opt
```

```
$D
[1] 0.364241 # La racine kième de la variance généralisée

$A
[1] 3.485722 # La variance moyenne des coefficients

$Ge
[1] 0.486 # La variance de la variance normalisée minimax sur X

$Dea
[1] 0.348 # Limite inférieure de l'efficacité D pour les conceptions théoriques approximatives.
Il est égal à exp(1-1/Ge).

# les 14 expériences à réaliser
$design
     tfo va hb nb tpb tfu Hr
92     2  2  2  2   1   3  2
220    1  2  1  2   1   2  4
500    2  1  1  1   2   2  3
515    2  2  2  3   2   4  4
516    2  1  1  3   1   4  1
761    1  1  2  1   3   3  4
905    1  1  2  3   3   2  2
928    2  2  2  1   3   2  1
934    1  2  1  2   3   4  3
965    1  1  2  2   2   1  1
970    2  2  1  1   3   1  2
1023   1  2  2  3   1   1  3
1111   1  2  1  3   2   3  1
1127   1  2  2  1   1   4  2

# numéro de ligne des 14 essais prisent dans le plan complet
$rows
 [1]   92  220  500  515  516  761  905  928  934  965  970 1023 1111 1127
```

```
# Vérification de la qualité du plan
options(contrasts=c("contr.sum", "contr.sum"))
Xopt <- model.matrix(~., plan.opt$design)
round(solve(t(Xopt)%*%Xopt),1) # on a arrondi le plan à 1 chiffre après virgule
```

```
            (Intercept) tfo1  va1  hb1 nb.L nb.Q tpb.L tpb.Q tfu.L tfu.Q tfu.C Hr.L Hr.Q Hr.C
(Intercept)         0.1  0.0  0.0  0.0  0.0  0.0   0.0   0.0   0.0   0.0   0.0  0.0  0.0  0.0
tfo1                0.0  0.1  0.0  0.0  0.0  0.0   0.0   0.0   0.0   0.0   0.0  0.0  0.0  0.0
va1                 0.0  0.0  0.1  0.0  0.0  0.0   0.0   0.0  -0.1   0.0   0.0 -0.1  0.0  0.0
hb1                 0.0  0.0  0.0  0.1 -0.1  0.0   0.0   0.0   0.0   0.0   0.0 -0.1  0.0  0.0
nb.L                0.0  0.0  0.0 -0.1  0.3  0.0   0.0   0.0   0.0  -0.1   0.0  0.0  0.0  0.0
nb.Q                0.0  0.0  0.0  0.0  0.0  0.3   0.0   0.0   0.0   0.0   0.0  0.0  0.0  0.0
tpb.L               0.0  0.0  0.0  0.0  0.0  0.0   0.3   0.0   0.0   0.0   0.0  0.0  0.0  0.0
tpb.Q               0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.2   0.0  -0.1   0.0  0.0  0.0  0.0
tfu.L               0.0  0.0 -0.1  0.0  0.0  0.0   0.0   0.0   0.3   0.0   0.0  0.0  0.1  0.0
tfu.Q               0.0  0.0  0.0  0.0 -0.1  0.0   0.0  -0.1   0.0   0.4   0.0 -0.1  0.0  0.0
tfu.C               0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0   0.0   0.0   0.4 -0.1  0.0 -0.1
Hr.L                0.0  0.0 -0.1 -0.1  0.0  0.0   0.0   0.0   0.0  -0.1  -0.1  0.4  0.0  0.0
Hr.Q                0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0   0.1   0.0   0.0  0.0  0.3  0.1
Hr.C                0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0   0.0   0.0  -0.1  0.0  0.1  0.3
```

On a une pseudo-orthogonalité (presque diagonale), il y a certains effets principaux non indépendants d'autres avec des covariances de l'ordre de 0.1 (parfois négative).

4. On veut réaliser le plan orthogonal

On utilise 144 essais pour avoir un plan orthogonal

```
plan.orth <- optFederov(~tfo+va+hb+nb+tpb+tfu+Hr, plan.ex2, nTrials = 144,  criterion = "D")

Xorth <- model.matrix(~tfo+va+hb+nb+tpb+tfu+Hr, plan.orth$design)
round(solve(t(Xorth)%*%Xorth),2)
```

```
            (Intercept) tfo1  va1  hb1 nb.L nb.Q tpb.L tpb.Q tfu.L tfu.Q tfu.C Hr.L Hr.Q Hr.C
(Intercept)        0.01 0.00 0.00 0.00 0.00 0.00  0.00  0.00  0.00  0.00  0.00 0.00 0.00 0.00
tfo1               0.00 0.01 0.00 0.00 0.00 0.00  0.00  0.00  0.00  0.00  0.00 0.00 0.00 0.00
va1                0.00 0.00 0.01 0.00 0.00 0.00  0.00  0.00  0.00  0.00  0.00 0.00 0.00 0.00
hb1                0.00 0.00 0.00 0.01 0.00 0.00  0.00  0.00  0.00  0.00  0.00 0.00 0.00 0.00
nb.L               0.00 0.00 0.00 0.00 0.02 0.00  0.00  0.00  0.00  0.00  0.00 0.00 0.00 0.00
nb.Q               0.00 0.00 0.00 0.00 0.00 0.02  0.00  0.00  0.00  0.00  0.00 0.00 0.00 0.00
tpb.L              0.00 0.00 0.00 0.00 0.00 0.00  0.02  0.00  0.00  0.00  0.00 0.00 0.00 0.00
```

```
tpb.Q               0.00 0.00 0.00 0.00 0.00 0.00  0.00  0.02  0.00  0.00  0.00 0.00 0.00 0.00
tfu.L               0.00 0.00 0.00 0.00 0.00 0.00  0.00  0.00  0.03  0.00  0.00 0.00 0.00 0.00
tfu.Q               0.00 0.00 0.00 0.00 0.00 0.00  0.00  0.00  0.00  0.03  0.00 0.00 0.00 0.00
tfu.C               0.00 0.00 0.00 0.00 0.00 0.00  0.00  0.00  0.00  0.00  0.03 0.00 0.00 0.00
Hr.L                0.00 0.00 0.00 0.00 0.00 0.00  0.00  0.00  0.00  0.00  0.00 0.03 0.00 0.00
Hr.Q                0.00 0.00 0.00 0.00 0.00 0.00  0.00  0.00  0.00  0.00  0.00 0.00 0.03 0.00
Hr.C                0.00 0.00 0.00 0.00 0.00 0.00  0.00  0.00  0.00  0.00  0.00 0.00 0.00 0.03
```

Le plan orthogonal est de bonne qualité (minimise bien la variance, avec indépendance des effets principaux), mais demande beaucoup d'expérience (144 essais).

Exercice 10. Création d'un nouveau parfum (exemple fictif, par Roxane Schott)

La situation présentée dans cet exercice est fictive.

Un parfumeur souhaite créer une nouvelle fragrance fraiche et estivale. La structure d'un parfum se partage en trois groupes : les notes de tête, en général des senteurs vertes, fraiches et volatiles ; les notes de cœur, souvent des senteurs fleuries ou épicées qui caractérisent le parfum ; et enfin les notes de fond, boisées le plus communément. Ce sont aussi ces dernières qui s'évaporent le plus lentement et font durer le parfum.

Plus la note de tête est importante, plus le résultat est frais. A l'inverse, plus les notes de fond sont présentes, et plus le parfum est riche et opulent.

Le parfumeur peut réaliser plusieurs compositions, dans le but de les faire par la suite analyser par des nez.

Les facteurs étudiés sont :

- La note de tête
  Trois modalités correspondant à des matières vertes et fraiches : bambou, figuier, feuille de violette
- La note de cœur
  Cinq modalités : fleur de cerisier, chèvrefeuille, aubépine, baies roses, cumin
- La note de fond
  Deux modalités : ambre, cyprès

L'objectif est de fournir au parfumeur un plan d'expérience permettant de choisir au mieux la composition la plus satisfaisante.

1. **A quel intervalle le nombre d'essais doivent-ils appartenir pour pouvoir tester l'ensemble des effets? Combien d'essais sont nécessaires pour avoir une orthogonalité entre les facteurs ? Ecrire les calculs.**

Nous avons ici un facteur à 2 modalités, un facteur à 3 modalités et un facteur à 5 modalités.

Le nombre maximal d'essais est 2*3*5= 30 essais (plan complet).

Au minimum, il faut autant d'essais que de degrés de liberté nécessaires pour estimer tous les effets. On somme les degrés de liberté pour la constante, et pour les facteurs à 2, 3 et 5 modalités. 1 + (2-1) + (3-1) + (5-1) = 8 essais. Il faut au moins huit essais car on a huit coefficients à estimer.

Le plan doit contenir entre 8 et 30 essais.

Pour avoir l'orthogonalité des facteurs, il faut que toutes les combinaisons de deux facteurs soient testées un même nombre de fois dans le plan d'expérience.

Pour que les facteurs 1 et 2 soient orthogonaux, il faut 3*5 = 15 essais. Pour que les facteurs 1 et 3 soient orthogonaux, il faut 3*2=6 essais. Enfin, pour que les facteurs 2 et 3 soient orthogonaux, il faut 5*2 = 10 essais.

N >= PPCM (6, 10, 15) = 30

L'orthogonalité des facteurs nécessite 30 essais, donc le plan complet est le plan permettant l'absence de confusion des effets.

2. **Construire sur R le plan $L_{30}235$ avec DoE.base. Commenter sa qualité.**

```
library(DoE.base)
plan <- oa.design(nlevels=c(3,5,2))
colnames(plan) <- c("Tête","Coeur","Fond")
   Tête Coeur Fond
1     3     1    2
```

```
2      2      2     2
3      1      1     1
4      2      4     2
5      1      4     1
6      3      5     1
7      1      3     1
8      3      3     2
9      1      5     2
10     3      2     2
11     2      4     1
12     1      2     2
13     1      3     2
14     2      2     1
15     3      4     2
16     2      3     2
17     2      5     2
18     2      1     2
19     1      5     1
20     3      4     1
21     3      3     1
22     3      1     1
23     2      1     1
24     2      5     1
25     2      3     1
26     1      2     1
27     3      5     2
28     1      4     2
29     3      2     1
30     1      1     2
class=design, type= full factorial
```

Test de la qualité du plan : Le modèle considéré est un modèle linéaire considérant également les effets d'interactions, ce qui est cohérent dans le contexte d'une association d'odeurs.

```
options(contrasts = c("contr.sum", "contr.sum"))
X <- model.matrix(~Tête+Coeur+Fond+Tête:Coeur+Tête:Fond+Coeur:Fond,data=plan)
solve(t(X)%*%X)
```

Il s'agit ici de s'intéresser aux covariances entre facteurs, entre interactions et facteurs, et entre interactions. La matrice ($X'X^{-1}$) est presque diagonale, et les covariances entre tous les facteurs ainsi que les interactions sont très faibles (comprises entre 10^{-17} et 10^{-34}). Les différents effets linéaires et d'interaction pourront être estimés sans confusion avec un tel plan d'expérience.

3. **Le parfumeur vous informe que le nombre d'essais que vous proposez est trop élevé. Seuls 20 tests pourront être réalisés, les moyens étant limités. Proposer un plan d'expérience obtenu sur R, adapté pour optimiser l'analyse des effets des différents facteurs.**

La création d'un plan optimal est réalisée grâce à la fonction optFederov de la librairie AlgDesign, avec 20 essais au lieu des 30 essais du plan complet.

```
library(AlgDesign)
set.seed(456)
planopt<-optFederov(~Tête+Coeur+Fond, data=plan, nTrials=20,eval=TRUE)
$design
   Tête Coeur Fond
1      2      1     1
2      2      2     1
3      1      4     1
4      1      4     2
6      3      5     2
7      1      2     2
8      3      1     2
10     1      1     1
12     2      5     2
14     2      2     2
15     2      3     1
16     2      3     2
17     3      5     1
20     3      3     1
23     2      4     1
24     2      1     2
25     3      2     1
26     1      5     1
29     1      3     2
30     3      4     2
```

La qualité du plan doit être testée. Si l'utilisation d'un plan optimisé en 20 essais au lieu de 30 fait perdre en orthogonalité entre les facteurs, il est nécessaire de vérifier que les effets pourront tout de même être étudiés avec une bonne qualité.

```
options(contrasts = c("contr.sum", "contr.sum"))
X2 <- model.matrix(~Tête+Coeur+Fond+Tête:Coeur+Tête:Fond+Coeur:Fond,
data=planopt$design)
round(t(X2)%*%X2,3)
               (Intercept) Tête.L Tête.Q Coeur.L Coeur.Q Coeur.C Coeur^4  Fond1 Tête.L:Coeur.L Tête.Q:Coeur.L Tête.L:Coeur.Q Tête.Q:Coeur.Q Tête.L:Coeur.C
(Intercept)         20.000  0.000 -1.633   0.000   0.000   0.000   0.000  0.000          0.224          1.162          0.567          0.327          0.671
Tête.L               0.000  6.000  0.000   0.224   0.567   0.671   0.423  0.000          0.474          0.091          0.134          0.231         -0.158
Tête.Q              -1.633  0.000  7.333   1.162   0.327  -0.387  -0.439  0.000          0.091         -0.474          0.231         -0.134          0.274
Coeur.L              0.000  0.224  1.162   4.000   0.000   0.000   0.000  0.000          0.212         -0.204          0.299          0.311          0.283
Coeur.Q              0.000  0.567  0.327   0.000   4.000   0.000   0.000  0.000          0.299          0.311          0.152         -0.379          0.000
Coeur.C              0.000  0.671 -0.387   0.000   0.000   4.000   0.000  0.000          0.283          0.000          0.000          0.414         -0.212
Coeur^4              0.000  0.423 -0.439   0.000   0.000   0.000   4.000  0.000          0.160         -0.093         -0.045          0.235         -0.187
Fond1                0.000  0.000  0.000   0.000   0.000   0.000   0.000 20.000          0.224         -0.387          2.079         -0.982         -1.565
Tête.L:Coeur.L       0.224  0.474  0.091   0.212   0.299   0.283   0.160  0.224          1.250          0.087          0.127          0.122          0.000
Tête.Q:Coeur.L       1.162  0.091 -0.474  -0.204   0.311   0.000  -0.093 -0.387          0.087          1.417          0.122         -0.127          0.115
Tête.L:Coeur.Q       0.567  0.134  0.231   0.299   0.152   0.000  -0.045  2.079          0.127          0.122          1.179          0.062          0.169
Tête.Q:Coeur.Q       0.327  0.231 -0.134   0.311  -0.379   0.414   0.235 -0.982          0.122         -0.127          0.062          1.488          0.000
Tête.L:Coeur.C       0.671 -0.158  0.274   0.283   0.000  -0.212  -0.187 -1.565          0.000          0.115          0.169          0.000          1.250
Tête.Q:Coeur.C      -0.387  0.274  0.158   0.000   0.414  -0.204   0.417 -1.162          0.115          0.000          0.000         -0.169         -0.087
Tête.L:Coeur^4       0.423 -0.179  0.173   0.160  -0.045  -0.187  -0.152 -0.423         -0.038          0.065          0.096         -0.018          0.170
Tête.Q:Coeur^4      -0.439  0.173  0.179  -0.093   0.235   0.417  -0.519 -0.732          0.065          0.038         -0.018         -0.096         -0.076
Tête.L:Fond1         0.000  0.000  0.000   0.224   2.079  -1.565  -0.423  0.000         -0.158          0.091         -0.401          0.849         -0.474
Tête.Q:Fond1         0.000  0.000  0.000  -0.387  -0.982  -1.162  -0.732 -1.633          0.091          0.158          0.849          0.401         -0.639
Coeur.L:Fond1        0.000  0.224 -0.387   0.000   0.000   0.000   0.000  0.000          0.778         -0.367         -0.418         -0.518          0.566
Coeur.Q:Fond1        0.000  2.079 -0.982   0.000   0.000   0.000   0.000  0.000         -0.418         -0.518          0.152         -0.262          0.239
Coeur.C:Fond1        0.000 -1.565 -1.162   0.000   0.000   0.000   0.000  0.000          0.566         -0.490          0.239          0.000         -0.071
Coeur^4:Fond1        0.000 -0.423 -0.732   0.000   0.000   0.000   0.000  0.000         -0.374         -0.278          0.587          0.078          0.615
               Tête.Q:Coeur.C Tête.L:Coeur^4 Tête.Q:Coeur^4 Tête.L:Fond1 Tête.Q:Fond1 Coeur.L:Fond1 Coeur.Q:Fond1 Coeur.C:Fond1 Coeur^4:Fond1
(Intercept)            -0.387          0.423         -0.439        0.000        0.000         0.000         0.000         0.000         0.000
Tête.L                  0.274         -0.179          0.173        0.000        0.000         0.224         2.079        -1.565        -0.423
Tête.Q                  0.158          0.173          0.179        0.000        0.000        -0.387        -0.982        -1.162        -0.732
Coeur.L                 0.000          0.160         -0.093        0.224       -0.387         0.000         0.000         0.000         0.000
Coeur.Q                 0.414         -0.045          0.235        2.079       -0.982         0.000         0.000         0.000         0.000
Coeur.C                -0.204         -0.187          0.417       -1.565       -1.162         0.000         0.000         0.000         0.000
Coeur^4                 0.417         -0.152         -0.519       -0.423       -0.732         0.000         0.000         0.000         0.000
Fond1                  -1.162         -0.423         -0.732        0.000       -1.633         0.000         0.000         0.000         0.000
Tête.L:Coeur.L          0.115         -0.038          0.065       -0.158        0.091         0.778        -0.418         0.566        -0.374
Tête.Q:Coeur.L          0.000          0.065          0.038        0.091        0.158        -0.367        -0.518        -0.490        -0.278
Tête.L:Coeur.Q          0.000          0.096         -0.018       -0.401        0.849        -0.418         0.152         0.239         0.587
Tête.Q:Coeur.Q         -0.169         -0.018         -0.096        0.849        0.401        -0.518        -0.262         0.000         0.078
Tête.L:Coeur.C         -0.087          0.170         -0.076       -0.474       -0.639         0.566         0.239        -0.071         0.615
Tête.Q:Coeur.C          1.417         -0.076         -0.170       -0.639        0.474        -0.490         0.000         0.367         0.324
Tête.L:Coeur^4         -0.076          1.121         -0.062       -0.299       -0.173        -0.374         0.587         0.615        -0.859
Tête.Q:Coeur^4         -0.170         -0.062          1.545       -0.173        0.299        -0.278         0.078         0.324         0.262
Tête.L:Fond1           -0.639         -0.299         -0.173        6.000        0.000         0.224         0.567         0.671         0.423
Tête.Q:Fond1            0.474         -0.173          0.299        0.000        7.333         1.162         0.327        -0.387        -0.439
Coeur.L:Fond1          -0.490         -0.374         -0.278        0.224        1.162         4.000         0.000         0.000         0.000
Coeur.Q:Fond1           0.000          0.587          0.078        0.567        0.327         0.000         4.000         0.000         0.000
Coeur.C:Fond1           0.367          0.615          0.324        0.671       -0.387         0.000         0.000         4.000         0.000
Coeur^4:Fond1           0.324         -0.859          0.262        0.423       -0.439         0.000         0.000         0.000         4.000
```

Certaines valeurs de la matrice ($X'X^{-1}$) sont supérieures à celles obtenues avec le plan complet, ce qui est totalement attendu puisque le plan complet comportait le nombre d'essais nécessaires pour obtenir l'orthogonalité des facteurs. Cependant, la matrice est presque diagonale, car les valeurs en dehors de la diagonale sont proches de 0.

Ainsi, bien que n'assurant pas l'orthogonalité parfaite entre les facteurs, ce plan permet une assez bonne distinction des effets, vu les contraintes techniques imposées.

Exercice 11. Maximiser le rendement d'une prairie en trouvant le mix d'espèces optimal (contexte réel, par AV. de Croutte)

On cherche maintenant à jouer sur la composition de la pâture, facteur pouvant aussi influencer le rendement. On considère que les graines sont mélangées avant d'être semées, on a donc une homogénéité de composition au sein d'une même parcelle. On dispose de 4 espèces : Ray-grass, Plantain, Trèfle blanc et Trèfle violet. Pour une question pratique au niveau des créations des mélanges, on va considérer que les espèces ne peuvent prendre que les proportions suivantes : 0, ¼, ½, ¾, ou 1. Les proportions sont considérées en proportions du poids total du mix de graines, avec un poids total toujours fixe. Le Plantain aide au développement des autres espèces mais est moins apprécié des moutons. On limitera sa participation à ¼ de la pâture maximum.

De plus, on dispose de 2 sites pour planter les parcelles : Ashley Dene et Iversen. Il faudra éviter qu'il y ait un biais à cause de cette différence et l'ajouter comme facteur à 2 modalités.

1. **Quel problème se pose pour l'élaboration du plan et comment y remédier ?**

Nous avons 5 proportions possibles pour chaque espèce, et 2 pour le Plantain (0 ou ¼), or il faut que la somme fasse toujours 1, mais on ne peut pas ajouter de condition globale sur les facteurs. On peut donc étudier cela sous la forme d'un plan avec 1 facteurs à 4 modalités, et 3 facteurs à 3 modalités en abordant le problème par les quarts de mix de graines et non par les espèces. On aura donc les facteurs suivants avec leurs modalités :

- 1er quart : Ray-grass, Plantain, Trèfle blanc ou Trèfle violet,
- 2^{e} quart : Ray-grass, Trèfle blanc ou Trèfle violet
- 3^{e} quart : Ray-grass, Trèfle blanc ou Trèfle violet,
- 4^{e} quart : Ray-grass, Trèfle blanc ou Trèfle violet.

On ajoute à ces facteurs le facteur lieu à 2 modalités :

- Lieu : Ashley Dene ou Iversen.

2. Quel plan construire avec ces 5 facteurs pour qu'ils soient orthogonaux ?

On a 5 facteurs, dont 1 à 4 modalités, 3 à 3 modalités et 1 à 2 modalités, on peut donc construire un **plan asymétrique**. On devra avoir au minimal 1 + 1*(2-1) + 3*(3-1) + 1*(4-1) = 11 essais pour pouvoir estimer les facteurs principaux.

On peut faire au total 216 essais différents avec ces facteurs ($2*3^3*4$), et pour construire un plan où les facteurs sont orthogonaux, il faut PPCM(8,9,12,16) = 144 essais. Ainsi, on n'est pas obligé d'avoir le plan complet pour avoir l'orthogonalité des facteurs, mais 48 reste un nombre très élevé d'essais.

```
library(AlgDesign)
comp <-  gen.factorial(levels=c(4,3,3,3,2),nVars=5,varNames=c("premquart",
"deuxquart","troisquart","quatrequart","lieu"))
dim(comp)
[1] 216   5
```

On a bien 216 essais pour le plan complet.

3. Chaque lieu a une surface limitée et ne peut contenir que 15 parcelles. Quel plan construire alors ?

On peut chercher un plan optimal avec l'algorithme de Federov sur R. Cela baissera la qualité du plan, mais il n'y a pas de choix possible si on ne veut que 30 essais. On utilise nRep=100 dans la fonction de l'algorithme de Federov pour s'assurer de ne pas tomber sur un minimum local.

```
plan_opt <- optFederov(~.,data=comp,30,nRep=100)
plan_opt$design

    premquart deuxquart troisquart quatrequart lieu
3           1        -1         -1          -1   -1
5          -3         0         -1          -1   -1
30         -1         0          1          -1   -1
33         -3         1          1          -1   -1
39          1        -1         -1           0   -1
46         -1         1         -1           0   -1
52          3        -1          0           0   -1
57         -3         1          0           0   -1
64          3        -1          1           0   -1
80          3         0         -1           1   -1
91          1         0          0           1   -1
95          1         1          0           1   -1
96          3         1          0           1   -1
98         -1        -1          1           1   -1
116         3         0         -1          -1    1
122        -1        -1          0          -1    1
123         1        -1          0          -1    1
132         3         1          0          -1    1
144         3         1          1          -1    1
148         3        -1         -1           0    1
153        -3         1         -1           0    1
161        -3         0          0           0    1
162        -1         0          0           0    1
175         1         0          1           0    1
179         1         1          1           0    1
181        -3        -1         -1           1    1
190        -1         1         -1           1    1
191         1         1         -1           1    1
205        -3        -1          1           1    1
212         3         0          1           1    1
```

4. Quelle est la qualité du nouveau plan ? Que peut-on en dire ?

```
X <- model.matrix(~.,data=plan_opt$design)
round(solve(t(X)%*%X),3)

          (Intercept) premquart-1 premquart1 premquart3 deuxquart0 deuxquart1
(Intercept)     0.350      -0.147     -0.146     -0.147     -0.102     -0.105
premquart-1    -0.147       0.314      0.146      0.147      0.001      0.014
premquart1     -0.146       0.146      0.272      0.145      0.014      0.016
premquart3     -0.147       0.147      0.145      0.258      0.001      0.015
deuxquart0     -0.102       0.001      0.014      0.001      0.214      0.102
deuxquart1     -0.105       0.014      0.016      0.015      0.102      0.194
troisquart0    -0.077      -0.014     -0.016     -0.015     -0.011     -0.012
troisquart1    -0.087      -0.014     -0.002     -0.014     -0.009     -0.001
quatrequart0   -0.125       0.014      0.002      0.014      0.009      0.001
quatrequart1   -0.110      -0.001     -0.014     -0.001     -0.002     -0.011
lieu           -0.004       0.004      0.004      0.001     -0.004     -0.003

            troisquart0 troisquart1 quatrequart0 quatrequart1   lieu
(Intercept)      -0.077      -0.087       -0.125       -0.110 -0.004
premquart-1      -0.014      -0.014        0.014       -0.001  0.004
premquart1       -0.016      -0.002        0.002       -0.014  0.004
premquart3       -0.015      -0.014        0.014       -0.001  0.001
deuxquart0       -0.011      -0.009        0.009       -0.002 -0.004
```

```
deuxquart1         -0.012       -0.001       0.001       -0.011 -0.003
troisquart0         0.194        0.092      -0.001        0.011  0.003
troisquart1         0.092        0.204       0.008        0.009  0.000
quatrequart0       -0.001        0.008       0.204        0.112  0.000
quatrequart1        0.011        0.009       0.112        0.214  0.004
lieu                0.003        0.000       0.000        0.004  0.034
```

La variance pour la variable lieu est assez faible mais les autres sont plutôt importantes. De plus, il y a de nombreuses covariances (inférieures aux variances mais tout de mêmes présentes) ce qui indique des confusions entre la plupart des variables. Le plan est donc maintenant très loin d'être orthogonal, et de piètre qualité.

Pour mener à bien cette expérience, peut-être faudrait-il trouver un 3^{e} lieu pour pouvoir augmenter le nombre d'essais, malgré une augmentation aussi du nombre de modalités du facteur lieu. On pourrait aussi réduire la taille de parcelles pour en faire rentrer plus sur les 2 sites, mais on prend alors le risque d'un effet bord prononcé sur le rendement.

Exercice 12. Culture du blé (exemple fictif, par Aurélien Carnet)

Un fermier à décider de modifier la production de blé sur ces parcelles. Il aimerait que sa production soit la plus importante. Pour cela, il a décidé de faire jouer plusieurs paramètres : la variété du blé (Rubisko, Cellule et Apache), la dose d'engrais (1,2,3,4), la dose d'eau (1,2,3,4), la dose de pesticide (1,2) et la dose de désherbant (1,2).

Soit 2 facteurs à 2 modalités, 1 facteur à 3 modalités et 2 facteurs à 4 modalités.

1. **Combien faudrait-il faire d'essais pour avoir le plan complet ?**

On multiplie le nombre de modalités

2*2*3**4*4= 192

Pour avoir le plan complet, il faudrait faire 192 essais

2. **Combien faut-il avoir d'essais ?**

Nombre minimum d'essais : 1+2*(2-1) +(3-1)+2*(4-1)=11 essais

3. **Construire le plan complet**

```
set.seed(123)
library(DoE.base)
plancompl <- fac.design(nlevels=c(2,2,3,4,4), factor.names= c("désherbant","insecticide",
"variétés","eau", "engrais"))
creating full factorial with 192 runs ...
    désherbant insecticide variétés eau engrais
1            1           2        1   2       4
2            1           2        3   3       4
3            2           1        1   2       1
4            2           1        1   3       4
5            2           1        1   1       2
6            2           1        3   2       3
7            1           2        2   4       1
8            2           1        3   4       4
9            1           2        2   4       4
10           1           1        3   1       4
11           2           1        2   4       2
[ . . . ]
188          1           2        1   3       2
189          2           2        2   3       2
190          1           2        3   4       4
191          1           2        2   3       3
192          1           2        1   2       1
class=design, type= full factorial
```

4. **Construire le plan optimal en 24 essais**

```
library(AlgDesign)
plan.2.Dopt<-optFederov(~.,plancompl,nTrials=24,criterion="D")
plan.2.Dopt
$D
[1] 0.3790187

$A
[1] 3.101372

$Ge
[1] 0.875

$Dea
```

```
[1] 0.867

$design
   désherbant insecticide variétés eau engrais
6           2           1        3   2       3
18          2           1        1   3       1
20          1           1        3   4       4
25          2           1        2   4       4
38          1           2        1   4       3
39          2           2        3   1       4
46          1           1        1   1       3
53          2           2        1   2       1
81          1           2        2   3       1
84          2           2        1   2       2
93          1           1        1   4       2
104         1           1        2   2       1
107         2           1        2   1       2
111         1           1        2   2       4
116         2           1        3   3       2
121         2           2        2   4       3
122         1           2        3   3       2
125         1           2        2   1       2
130         1           1        3   1       1
133         2           2        3   4       1
145         1           2        3   2       3
163         2           2        1   1       4
173         1           2        1   3       4
186         2           1        2   3       3

$rows
 [1]   6  18  20  25  38  39  46  53  81  84  93 104 107 111 116 121 122 125 130 133 145 163 173
[24] 186
```

Exercice 13. Etude de préférences

Une marque veut lancer une nouvelle gamme de produit et cherche un logo qui sera présent sur tous ses emballages. Elle hésite pour plusieurs critères et souhaite donc avoir l'avis des consommateurs. IL y a 4 couleurs, 4 formes, 2 taille de police, 3 style de police et 3 symboles (fleur, feuille, arble) possible sur le logo. Par soucis économique l'entreprise ne peut pas tester l'ensemble des combinaisons et souhaiterait faire noter uniquement 10 produits .

1. **Combien d'essais faudrait-il pour avoir le plan complet ? Combien faut-il d'essais au minimum si on veutl'orthogonalité entre les facteurs 2 à 2 ?**

plan complet= 2*3^{2}*4^{2}= 288 essais

Pour avoir l'orthogonalité il faut calculer le PPCM des produits des facteurs pris 2 à 2. On a alors PPCM($2*3,2*4,3*4,4*4,3*3$) = PPCM(6,8,12,16,9)= 144

2. **Construire le plan complet puis le plan optimisé avec seulement 10 essais.**

```
library(DoE.base)

data1 = fac.design(nlevels=c(4,4,3,3,2),factor.names= c("couleur","forme","style","symbole", "ta
ille"))

X = model.matrix(~.,data=data1)

library(AlgDesign)
data2 = gen.factorial(levels=c(4,4,3,3,2),nVars=5,varNames=c("couleur","forme","style","symbole"
, "taille"))
plan.opt = optFederov(~.,data2,nTrials=10,nRepeats=20)
plan.opt

 $D
 [1] 2.005319

 $A
 [1] 0.7539683

 $Ge
 [1] 0.84

 $Dea
 [1] 0.827

 $design
    couleur forme style symbole taille
```

```
16         3     3    -1      -1     -1
36         3    -3     1      -1     -1
109       -3     3    -1       1     -1
129       -3    -3     1       1     -1
132        3    -3     1       1     -1
145       -3    -3    -1      -1      1
189       -3     3     1      -1      1
244        3    -3    -1       1      1
256        3     3    -1       1      1
288        3     3     1       1      1

$rows
 [1]  16  36 109 129 132 145 189 244 256 288
```

3. **Comparer la qualité des plans en calculant le déterminant et commentez l'orthogonalité du plan optimisé.**

```
X = model.matrix(~.,data=data1)
a = solve(t(X)%*%X)
b = det(a)
b

 [1] 1.018908e-24

options(contrasts = c("contr.sum", "contr.sum"))
X1 = model.matrix(~.,data=plan.opt$design)
a1 = solve(t(X1)%*%X1)
b1 = det(a1)
b1

 [1] 1.537797e-08
```

Ici on a bien le déterminant du plan complet qui est beaucoup plus petit que celui du plan optimisé, il est donc logiquement plus performant que le plan optimisé

```
a1

             (Intercept)      couleur        forme       style      symbole       taille
(Intercept)  0.107142857 -0.005952381  0.000000000 0.000000000 -0.017857143  0.000000000
couleur     -0.005952381  0.011904762  0.000000000 0.000000000 -0.005952381  0.000000000
forme        0.000000000  0.000000000  0.011904762 0.005952381  0.000000000 -0.005952381
style        0.000000000  0.000000000  0.005952381 0.107142857  0.000000000  0.017857143
symbole     -0.017857143 -0.005952381  0.000000000 0.000000000  0.107142857  0.000000000
taille       0.000000000  0.000000000 -0.005952381 0.017857143  0.000000000  0.107142857
```

Ici le plan optimisé est plutôt bon, les valeurs dans la diagonale sont bien supérieures aux autres. les effets dépendent très faiblement de certains autres effets. Par exemple taille dépent très légèrement de couleur et forme.

4. **L'entreprise est prête à rajouter 4 tests si besoin, le développement des 10 premiers logos a couté moins cher que prévu. recalculez le déterminant et la matrice des effets en gardant les tests déjà présents. Commentez**

```
plan.opt2 <- optFederov(~.,data2,nTrials=14,augment=TRUE,rows=plan.opt$rows, nRepeats = 20)

X2 = model.matrix(~.,data=plan.opt2$design)
a2 = solve(t(X2)%*%X2)
b2 = det(a2)
b2

 [1] 1.883801e-09
```

Ici le déterminant est 10 fois plus faible que pour le premier plan, le plan est plus performant ce qui est logique car on fait plus d'essais. Mais comparé au déterminant du plan complet, il reste beaucoup plus grand. on peut dire que les déterminants des deux plans sont du même ordre de grandeur par rapport à celui du plan complet

```
a2

             (Intercept)      couleur        forme        style      symbole       taille
(Intercept)  0.075000000  0.004166667  0.000000000  0.000000000 -0.012500000  0.000000000
couleur      0.004166667  0.008333333  0.000000000  0.000000000 -0.004166667  0.000000000
forme        0.000000000  0.000000000  0.008333333  0.004166667  0.000000000 -0.004166667
style        0.000000000  0.000000000  0.004166667  0.075000000  0.000000000 -0.012500000
symbole     -0.012500000 -0.004166667  0.000000000  0.000000000  0.075000000  0.000000000
taille       0.000000000  0.000000000 -0.004166667 -0.012500000  0.000000000  0.075000000
```

Au vu de la matrice, les variables sont moins liées globalement les unes aux autres que lors du précédent test. Cependant cette diminution de dépendance reste très faible.

Donc au vu des résultats, l'ajout de 4 essais ne modifie l'orthogonalité que très faiblement. Connaissant le prix de développement d'un logo il est préférable pour la direction de ne pas rajouter les essais supplémentaires. En effet le premier plan avec les 10 essais garantie déjà un plan de très bonne qualité.

5. **Quelle autre solution l'entreprise pourrait choisir pour garantir une meilleure orthogonalité entre les facteurs sans devoir rajouter de tests?**

Diminuer le nombre de modalités des facteurs. Car si le nombre d'essais pour obtenir l'orthogonalité diminue (PPCM) alors avec seulement 10 essais on pourra obtenir un plan plus performant que précédemment. Par exemple si on enlève une modalité à un facteur qui en contient 4 et une à un facteur qui en contient 3, le PPCM obtenu est égal à 72. Cela veut qu'il ne faut plus que 72 essais pour obtenir l'orthogonalité. Donc avec seulement 10 essais la performance du plan serait meilleure que précédemment.

6. **On a fait noté les 10 essais choisis à un panel de 50 consommateurs. Quelle méthode doit être utilisé pour analyser les résultats?**

Pour analyser les résultats de ce plan, nous devrons effectuer une analyse de variance à 6 facteurs (1 facteurs à 2 modalités, 2 à 3 modalités et 2 à 4 modalités) ainsi qu'un facteur à 50 modalités (le facteur consommateur).

$$Yijklmn = \mu_i + \alpha_i + \beta_j + \gamma_k + \delta_l + \zeta_m + \eta_n + \varepsilon_{ijklmn}$$

Exercices divers

Exercice 1. La pâte à pizza (exemple fictif, par L. Sainton)

Partie 1 : Etude du collant d'une pate a pizza, Plan fractionnaire

Pour une compétition du meilleur pizzaiolo du monde, Ludovico souhaite fabriquer une pâte à pizza peu collante afin de cuisiner la pizza qui lui permettra d'obtenir le titre de maitre pizzaiolo.

Dans un premier temps Ludovico souhaite déterminer quels paramètres influencent le collant de sa pâte et ne souhaite pas multiplier les expériences pour une question de temps et de budget. Il ne souhaite pas réaliser plus de 8 pâtes à pizza.

D'après son expérience il sait que les ingrédients pour réaliser sa pâte peuvent varier entre :

- Farine (500 ou 550g)
- Eau (250 ou 280 ml)
- Levure (10 ou 20g)
- Temps de malaxage (5 ou 10 min)

1. **Quel plan d'expérience pourrait mettre en place Ludovico ?**

Ludovico pourrait réaliser un plan fractionnaire 2^{4-1}. En effet, il a 4 facteurs à deux modalités à étudier (farine, eau, levure et temps de malaxage) et ne peut réaliser plus de 8 expériences.

2. **Ecrire la matrice des essais, donner la liste des confusions et la résolution de ce plan.**

Farine = A ; Eau = B ; Levure = C ; Temps = D

<u>Première étape</u> : construction de la matrice des essais

Essai	A	B	C	AB	AC	BC	ABC
1	1	1	1	1	1	1	1
2	1	1	-1	1	-1	-1	-1
3	1	-1	1	-1	1	-1	-1
4	1	-1	-1	-1	-1	1	1
5	-1	1	1	-1	-1	1	-1
6	-1	1	-1	-1	1	-1	1
7	-1	-1	1	1	-1	-1	1
8	-1	-1	-1	1	1	1	-1

On choisit de confondre le facteur D avec l'interaction d'ordre 3 (ABC) car c'est la plus raisonnablement négligeable.

On obtient la matrice d'essai suivante :

Essai	A	B	C	D
1	1	1	1	1
2	1	1	-1	-1
3	1	-1	1	-1
4	1	-1	-1	1
5	-1	1	1	-1
6	-1	1	-1	1
7	-1	-1	1	1
8	-1	-1	-1	-1

<u>Liste des confusions :</u>

Le générateur d'aliase est ABCD.

Les confusions sont : A=BCD ; B=ACD ; C=ABD ; D=ABC ; AB=CD ; AC=BD ; BC=AD

La résolution du plan est IV, car les effets principaux sont confondus avec des interactions d'ordre 3.

3. **Avant de réaliser ses expériences Ludovico, vous demande conseil. Vous lui faites remarquer qu'il n'arrivera pas à trouver de conclusion optimale à son plan fractionnaire. Expliquer.**

En effet, le plan de Ludovico n'a que 7 degrés de liberté (8-1) ce qui lui permettrait de calculer seulement 7 paramètres il n'a donc pas assez pour calculer toutes les interactions d'ordre 2. Il pourrait donc manquer des interactions significatives.

4. **Vous finissez par convaincre Ludovico de réaliser un plan complet. Il obtient les valeurs suivantes. Y (220, 215, 300, 305, 200, 195, 307, 305, 140, 145, 275, 280, 138, 149, 280, 269). Analyser les résultats sur R.**

```
library (FactoMineR)
library(FrF2)
plan<-FrF2(nruns=16,nfactors=4,randomize = FALSE) # j'utilise randomize = false pour générer
l'expérience dans un ordre fixe pour faciliter la correspondance avec les Y donnés
Y<-c(220,215,300,305,200,195,307,305,140,145,275,280,138,149,280,269)
AovSum(Y~A+B+C+D+A:B+A:C+A:D+B:C+B:D+C:D,data=plan)
```

```
Ftest
             SS df    MS  F value    Pr(>F)
A             1  1     1   0.0085  0.930325
B         52785  1 52785 793.0150 1.057e-06 ***
C            86  1    86   1.2854  0.308305
D          8603  1  8603 129.2404 9.214e-05 ***
A:B           5  1     5   0.0761  0.793750
A:C          18  1    18   0.2714  0.624659
A:D          18  1    18   0.2714  0.624659
B:C          95  1    95   1.4282  0.285651
B:D        1314  1  1314  19.7418  0.006745 **
C:D          53  1    53   0.7897  0.414900
Residuals   333  5    67
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On remarque que les facteurs eau et temps de malaxage (B et D) ainsi que leur interaction ont un effet significatif sur l'effet collant de la pâte.

PARTIE 2 : Etude du collant d'une pâte à pizza par un plan pour surfaces de réponses (PAR L.SAINTON)

Remarque : Les résultats ne sont pas dans la continuité de la partie 1 pour des raisons de création de modèle interprétable.

Sachant maintenant que l'eau et le temps de malaxage sont déterminants pour influencer le collant de la pâte, Ludovico voudrait trouver les valeurs de ces paramètres qui lui permettrait d'avoir la pâte la moins collante possible. Il décide de faire varier la quantité d'eau entre 250 et 280 ml et le temps de malaxage entre 5 et 10 min.

5. **Proposer un plan optimal pour répondre à sa problématique. Combien d'expériences devra-t-il réaliser ?**

Le meilleur plan que pourrait effectuer Ludovico est un plan composite centré à 2 facteurs. Nombre d'expériences : $2^2+2*2+8 = 16$ expériences

6. **Construire le plan sur R et décrire à Ludovico 3 expériences qu'il va réaliser (une par partie du PCC : factoriel, étoiles, centre).**

```
library(rsm)
library(AlgDesign)
pcc <- ccd(2, coding=list (x1~(Eau-265)/15, x2~(Temp-7.5)/2.5), randomize = FALSE)
pcc
   run.order std.order      Eau      Temp Block
1          1         1 250.0000  5.000000     1
2          2         2 280.0000  5.000000     1
3          3         3 250.0000 10.000000     1
4          4         4 280.0000 10.000000     1
5          5         5 265.0000  7.500000     1
6          6         6 265.0000  7.500000     1
7          7         7 265.0000  7.500000     1
8          8         8 265.0000  7.500000     1
9          1         1 243.7868  7.500000     2
10         2         2 286.2132  7.500000     2
11         3         3 265.0000  3.964466     2
12         4         4 265.0000 11.035534     2
13         5         5 265.0000  7.500000     2
14         6         6 265.0000  7.500000     2
15         7         7 265.0000  7.500000     2
16         8         8 265.0000  7.500000     2

Data are stored in coded form using these coding formulas ...
x1 ~ (Eau - 265)/15
x2 ~ (Temp - 7.5)/2.5
```

Factoriel : 250 ml d'eau et 5 min
Etoilé : 265 ml d'eau et 3,9 min de malaxage
Centre : 265 ml d'eau et 7,5 min de malaxage

7. **Ludovico a réalisé ses 16 expériences avec le plan construit selon les lignes de code suivantes :**

```
library(rsm)
```

library(AlgDesign)
pcc <- ccd(2, coding=list (x1~(Eau-265)/15, x2~(Temp-7.5)/2.5), randomize = FALSE)

Voici ses résultats Y(220, 100, 300, 180, 160, 120, 150, 140, 274, 105, 113, 226, 160, 120, 150, 160). Analyser les sous R et déterminer les valeurs des paramètres pour avoir la pâte la moins collante.

Dans un premier temps on étudie la qualité du modèle quadratique avec la fonction rsm :

```
library(rsm)
library(AlgDesign)
pcc <- ccd(2, coding=list (x1~(Eau-265)/15, x2~(Temp-7.5)/2.5), randomize = FALSE)
Y<-c(220,100,300,180,160,120,150,140,274,105,113,226,160,120,150,160)
CR<- rsm(Y~SO(x1,x2), data=pcc)
summary(CR)

Call:
rsm(formula = Y ~ SO(x1, x2), data = pcc)

               Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  1.4500e+02  5.9588e+00  24.3336 3.131e-10 ***
x1          -5.9875e+01  5.9588e+00 -10.0481 1.521e-06 ***
x2           3.9976e+01  5.9588e+00   6.7086 5.307e-05 ***
x1:x2        2.0242e-14  8.4271e+00   0.0000 1.0000000
x1^2         2.7375e+01  5.9588e+00   4.5940 0.0009892 ***
x2^2         1.7375e+01  5.9588e+00   2.9158 0.0154101 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.9461,     Adjusted R-squared:  0.9192
F-statistic: 35.12 on 5 and 10 DF,  p-value: 4.975e-06

Analysis of Variance Table

Response: Y
            Df Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)   2  41465 20732.4 72.9854 1.083e-06
TWI(x1, x2)  1      0     0.0  0.0000  1.000000
PQ(x1, x2)   2   8410  4205.1 14.8035  0.001026
Residuals   10   2841   284.1
Lack of fit  3    841   280.2  0.9807  0.454554
Pure error   7   2000   285.7

Stationary point of response surface:
       x1        x2
 1.093612 -1.150382

Stationary point in original units:
       Eau       Temp
281.404181   4.624046

Eigenanalysis:
eigen() decomposition
$values
[1] 27.375 17.375

$vectors
            [,1]          [,2]
x1 -1.000000e+00  1.012124e-15
x2 -1.012124e-15 -1.000000e+00
```

On voit que la probabilité critique du test sur l'ajustement est supérieure à 0,05 (0,45), le modèle est donc correct et peut être conservé.

On peut maintenant regarder les effets significatifs et on remarque que les effets principaux sont significatifs tout comme l'effet quadratique des deux facteurs.

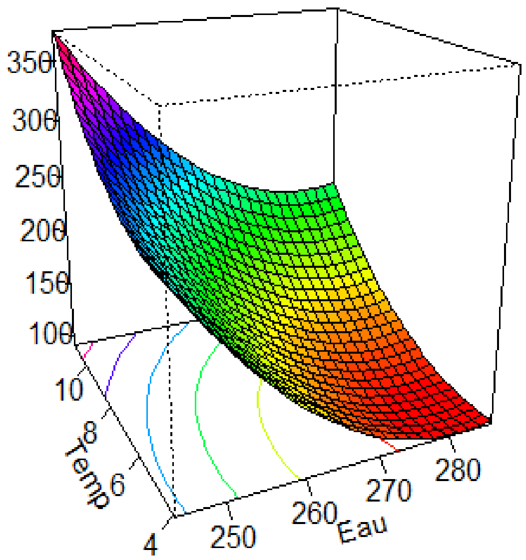
L'optimum est atteint lorsque les facteurs prennent les valeurs eau=281,4 ml et temps = 4,6 min.

Cet optimum correspond à un minimum car les valeurs propres sont du même signe et positives.

On peut construire les surfaces de réponse pour le visualiser :

```
library(rsm)
library(AlgDesign)
pcc <- ccd(2, coding=list (x1~(Eau-265)/15, x2~(Temp-7.5)/2.5), randomize = FALSE)
pcc
Y<-c(220,100,300,180,160,120,150,140,274,105,113,226,160,120,150,160)
CR<- rsm(Y~SO(x1,x2), data=pcc)
summary(CR)
```

```
persp(CR,~x1+x2,col=rainbow(50),contours="colors")
```



Pour conclure, Ludovico doit utiliser 281,4 ml d'eau et malaxer pendant 4,6 min afin d'obtenir la pâte la moins collante et gagner le concours.

Exercice 2. Création d'une nouvelle offre par Le Cercle (exemple réel, par A. Nowatski)

Le Cercle est une entreprise traiteur de plateaux repas pour entreprises. Elle s'est positionnée sur un marché déjà largement concurrencé mais pour se différencier, elle a choisi de proposer ses repas dans des conditionnements recyclés et réutilisables. Ils sont actuellement fabriqués en fibre de bambou, et très prochainement en épluchure de pomme de terre -plus écologique. Le seul emballage plastique subsistant est un plastique entourant les couverts, neufs de tout usage.

Les plateaux-repas sont livrés le matin même de la prestation -la veille dans des cas très particuliers-, dans des sacs isothermes et récupérés dans l'après-midi par les livreurs. Absolument tout est récupéré, du contenant au contenu délaissé. L'offre se compose de plateaux individuels, composés d'une entrée, un plat et un dessert et de buffets -appelés « cubes », ils sont composés d'étagères sur lesquelles sont disposées différents amuse-bouche, salés et sucrés-

Les produits sont de grande qualité, réalisés à partir de produits frais, de saison, et locaux autant que faire se peut. La démarche RSE et éco-responsable de l'entreprise est très aboutie et en représente l'ADN. Le Cercle est en constante recherche d'amélioration et d'innovation. La carte est changée à peu près 4 fois par an, notamment pour correspondre autant que possible à la saison. Le Cercle réfléchit à une nouvelle offre individuelle pour innover et rester dans la tendance. L'entreprise a plusieurs idées pour cette offre mais souhaiterait trouver LA combinaison parfaite.

Pour déterminer sa nouvelle offre, l'entreprise est prête à faire réaliser 16 plats différents, qui seraient des combinaisons de ses différentes idées.

Voici les idées dont il est question, en lien avec les tendances alimentaires actuelles :

- végétarienne : ni viande, ni poisson, ou avec l'un de ces produits
- sans gluten, ou avec gluten
- produits exclusivement français, ou possibilité de produits étrangers
- affichage des valeurs nutritionnelles ou non
- 0 emballage plastique, ou avec encore le film plastique autour des couverts
- sans lactose, ou avec
- repas diététique, ou « classique »
- repas froid, ou chaud

<u>1ère partie : choix des plats à réaliser (contexte réel)</u>

1. **Quel plan choisissez-vous de construire pour aider Le Cercle dans sa démarche?**

D'après les données, on a 8 facteurs qualitatifs à 2 modalités. On va donc utiliser un plan fractionnaire.
On sait également que l'on veut 16 tests, donc 16 essais.
On doit donc construire un plan 28-4.

2. Quels choix de confusions faites-vous et pourquoi ?

Nous devons exprimer 8 facteurs + la constante, soit 9 paramètres. Pour cela, nous ajoutons 4 facteurs au plan de base 2^4, qui permettra d'estimer les effets suivants :

I 1 2 3 4 12 13 14 23 24 34 123 124 134 234 1234

Nous choisissons de confondre les 4 facteurs ajoutés avec des interactions d'ordre 3

Si nous choisissons de confondre le facteur 5 avec l'interaction 123, le 6 avec 124, le 7 avec 134 et le 8 avec 234, nous obtenons :

I=1235=1246=1347=2348 comme générateurs d'alias initiaux.

On en déduit les produits de ces générateurs initiaux multipliés :

- 2 à 2 : I=3456=2457=1458=2367=1368=1278
- 3 à 3 : I=1567=4678=2568=3578
- Les 4 : 12345678

On a bien 2^4-1=15 générateurs d'alias.

3. Voici les confusions résultantes :

I	1	2	3	4	12	13	14	23	24	34	123	124	134	234	1234
1235	235	135	125	12345	35	25	2345	15	1345	1245	5	345	245	145	45
1246	246	146	12346	126	46	2346	26	1346	16	1236	346	6	236	136	36
1347	357	12347	1247	137	2347	47	37	1247	1237	17	47	237	7	127	27
2348	12348	348	148	238	348	248	1238	48	38	28	148	138	128	8	18

4. Quelle est la résolution de ce plan ?

Nous obtenons un plan de résolution IV -puisque la plus petite interaction avec laquelle un des effets principaux est confondue est une interaction d'ordre 3. Nous pouvons ainsi estimer les effets principaux sans ambiguïté, à partir du moment où l'on considère que les interactions d'ordre 3 ou plus sont négligeables. Les interactions d'ordre 2 pourront être estimées puisque supposées non nulles, mais elles ne pourront être estimées individuellement.

5. Construire le plan sous R et en vérifier la qualité.

Lignes de code :

```
library(FrF2)
plan <- FrF2(nruns=16,nfactors=8,factor.names=list(végétarien=c("oui","non"),
gluten= c("oui","non"),lactose=c("oui","non"), français=c("oui","non"), étiquette=
c("oui","non"),
emballage= c("oui","non"), froid= c("oui","non"), diététique= c("oui","non")),randomize=FALSE)
summary(plan)
```

```
Call:
FrF2(nruns = 16, nfactors = 8, factor.names = list(végétarien = c("oui",
    "non"), gluten = c("oui", "non"), lactose = c("oui",
    "non"), français = c("oui", "non"), étiquette = c("oui",
    "non"), emballage = c("oui", "non"), froid = c("oui",
    "non"), diététique = c("oui", "non")),
    randomize = FALSE)

Experimental design of type  FrF2
16  runs

Factor settings (scale ends):
  végétarien gluten lactose français étiquette emballage froid diététique
1        oui    oui     oui      oui       oui       oui   oui        oui
2        non    non     non      non       non       non   non        non

Design generating information:
$legend
[1] A=végétarien B=gluten     C=lactose    D=français   E=étiquette  F=emballage  G=froid
[8] H=diététique

$generators
[1] E=ABC F=ABD G=ACD H=BCD


Alias structure:
```

```
$fi2
[1] AB=CE=DF=GH AC=BE=DG=FH AD=BF=CG=EH AE=BC=DH=FG AF=BD=CH=EG AG=BH=CD=EF AH=BG=CF=DE

The design itself:
   végétarien gluten lactose français étiquette emballage froid diététique
1         oui    oui     oui      oui       oui       oui   oui        oui
2         non    oui     oui      oui       non       non   non        oui
3         oui    non     oui      oui       non       non   oui        non
4         non    non     oui      oui       oui       oui   non        non
5         oui    oui     non      oui       non       oui   non        non
6         non    oui     non      oui       oui       non   oui        non
7         oui    non     non      oui       oui       non   non        oui
8         non    non     non      oui       non       oui   oui        oui
9         oui    oui     oui      non       oui       non   non        non
10        non    oui     oui      non       non       oui   oui        non
11        oui    non     oui      non       non       oui   non        oui
12        non    non     oui      non       oui       non   oui        oui
13        oui    oui     non      non       non       non   oui        oui
14        non    oui     non      non       oui       oui   non        oui
15        oui    non     non      non       oui       oui   oui        non
16        non    non     non      non       non       non   non        non
class=design, type= FrF2

X<- model.matrix(~.,data=plan)
t(X)%*%X
            (Int) végétarien1 gluten1 lactose1 français1 étiquette1 emballage1 froid1 diététique1
(Intercept)    16           0       0        0         0          0          0      0           0
végétarien1     0          16       0        0         0          0          0      0           0
gluten1         0           0      16        0         0          0          0      0           0
lactose1        0           0       0       16         0          0          0      0           0
français1       0           0       0        0        16          0          0      0           0
étiquette1      0           0       0        0         0         16          0      0           0
emballage1      0           0       0        0         0          0         16      0           0
froid1          0           0       0        0         0          0          0     16           0
diététique1     0           0       0        0         0          0          0      0          16
```

Ce plan d'expérience va permettre au Cercle de tester plusieurs combinaisons de critères sans toutes les tester -8 !, ce qui fait tout de même beaucoup de possibilités… Le principe d'orthogonalité des plans fractionnaires va lui permettre de créer des combinaisons de caractéristiques équilibrées : chaque facteur et chaque combinaison de facteurs sera répétée autant de fois qu'un(e) autre.

La dernière commande R permet de vérifier la qualité du plan. Or la condition d'orthogonalité d'une expérience implique notamment que le produit X'X -que R a réalisé d'après la dernière commande-soit une matrice diagonale. Ce qui est ici le cas.

La condition d'orthogonalité simplifie considérablement l'analyse des résultats, dans la mesure où elle permet d'identifier sans problème les contributions des différents facteurs et éventuellement leurs différentes interactions, par analyse de la variance ou par régression multiple. Nous ne réaliserons pas ce type d'analyse ici car l'intérêt pour notre exemple n'est pas majeur, mais il est important de le préciser pour noter l'importance d'avoir un produit de matrice X'X diagonal.

2e partie : Réalisation des recettes

A présent, on s'intéresse aux facteurs influençant l'appréciation des 16 plats réalisés.

Le Cercle fait réaliser ses plats par 2 cuisines -appelées laboratoires- disctintes : Noon et Erisay. Chacune de ses cuisines dispose a priori des mêmes instructions, pourtant, des erreurs, différentes selon la cuisine, peuvent se produire durant la réalisation des plats.

Les plats vont être dégustés par les 4 directeurs de l'entreprise -toutes précautions sanitaires préservées, les 4 goûteront les mêmes 16 plats. Chacun a des goûts différents et donc une appricéation des plats différente les uns des autres.

Les plats seront présentés dans 2 contenants différents : la traditionnelle triologie en fibre de bambou -gris anthracite- ou le prototype de la trilogie en épluchures de pomme de terre -elle est verte et jusque là, certains l'ayant testés trouvaient qu'elle pouvait donner un goût au contenu.

L'entreprise responsable de la livraison -CBM- livre Le Cercle grâce à 4 de ses livreurs et tous ne sont pas aussi précautionneux les uns que les autres : les mouvements brusques sont susceptibles de déprécier l'aspect esthétique des plats à l'intérieur des contenants.

Enfin, la dégustation peut avoir lieu dans 4 pièces différentes du bureau, chacune ayant une ambiance différente des autres pour créer différents contextes de dégustation, correspondant aux situations que peuvent vivre les clients.

6. **Comment choisissez-vous d'étudier ces paramètres pour estimer leur influence sur la dégustation ?**

Les laboratoires et le contenant peuvent être vus comme 2 facteurs à 2 modalités.

Les 3 autres, que l'on appellera les facteurs « directeur », « livreur » et « pièce » pourront être vus comme des facteurs à 4 modalités.

On peut alors construire un plan $L_{16}4^3 2^2$ grâce à un plan 2^{8-4}. Dans ce plan, les facteurs à 4 modalités seront vus, un à un, comme 2 facteurs à 2 modalités. On aurait alors 8 facteurs à 2 modalités.

7. Le nombre d'essais réalisés est-il acceptable ?

Le nombre maximum d'essais à réaliser est 4^3*2^2=64.

Le nombre minimum d'essais à réaliser est lui de 1+3(4-1)+2(2-1)=12.

Or nous avons choisi de réaliser 16 essais donc le plan est acceptable.

8. Pourrions-nous étudier le facteur « heure » dans ce même modèle ? Aidez-vous de la résiduelle et du nombre de facteurs mis en œuvre.

La variance résiduelle du modèle est : 16 – (1+3+3+3+1+1)= 4 ddl.

D'après le ddl de la résiduelle, il serait donc possible d'ajouter un facteur à 1, 2, 3 ou 4 modalités. Si le facteur « heure » est à 3 modalités -on pourrait considérer matin/midi/après-midi par exemple-, alors nous pourrions l'estimer dans ce même plan et il resterait 2 ddl.

Cependant, nous avons déjà 8 facteurs dans notre modèle. Avec la constante, cela fait 9 paramètres à estimer. Si nous ajoutons un seul facteur à 2 modalités, le plan peut être de résolution IV si l'on confond les 5 facteurs ajoutés au plan de base à des interactions d'ordre 3 et 4. Si on ajoute plus d'un facteur à 2 modalités, il faut confondre des facteurs avec des interactions d'ordre 2 et se trouver en résolution III. Or en résolution III, on ne peut estimer que les facteurs principaux puisque toutes les interactions sont supposées négligeables. L'interprétation des résultats est donc à nuancer et à valider par des essais complémentaires pour confirmer ou non l'hypothèse des interactions négligeables.

L'intégration du facteur « heure » dans le modèle est donc :

- limité par la résiduelle : le facteur ne peut être à plus de 6 modalités et à 5 modalités, la résiduelle n'a plus de ddl
- limité par le nombre de facteurs déjà utilisé : au-delà de 2 modalités, il faut passer en résolution III et donc supposer que toutes les interactions sont négligeables

Par ailleurs, on peut supposer que l'heure à laquelle a lieu la dégustation peut avoir un effet sur les directeurs qui mangent -certaines personnes sont plus ou moins sensibles aux heures auxquelles elles mangent du sucré ou du salé par exemple. Introduire un tel facteur pourrait donc d'autant plus modifier la qualité du modèle qu'il pourrait avoir une interaction avec le facteur « directeur ».

Exercice 3. Étude des insectes pollinisateurs (inspiré d'un atelier de l'UE agroécologie réalisé en confinement, par Laurine Lambelin)

On souhaite effectuer des comptages des pollinisateurs sur les fleurs selon différents facteurs. La première partie porte sur le travail réalisé dans cet atelier : on s'intéresse à la diversité d'insectes en fonction du lieu, du type de plantes et de l'heure de la journée. La deuxième partie aurait pu être son prolongement mais n'a pas été réalisée. Elle vise à étudier les modalités de l'expérience (durée d'observation, localisation des surface d'échantillonnage…) afin d'optimiser le comptage d'insectes.

Les données de la première partie ont été adaptées pour l'exercice, celles de la deuxième partie sont fictives.

<u>PARTIE 1</u> : Recherche d'un plan optimal

Dans cette partie, on compte les pollinisateurs sur une surface d'un mètre carré (quadrat) selon 4 facteurs:

- Le lieu (parc / jardin)
- Le moment de la journée (9h / 12h / 15h / 18h)
- Le type de fleurs dans le quadrat (fleurs sauvages de la pelouse / églantine / marguerite / rose)
- La présence d'un arbre à proximité (oui / non)

1. Quel est le nombre d'essais associé au plan complet ? Combien faut-il d'essais au minimum si on veut que les facteurs soient orthogonaux 2 à 2 ?

Nb max d'essais = 2*4*4*2 = 64 pour le plan complet

Pour l'orthogonalité des facteurs il faut :

F1 ⊥ F4 donc minimum 2*2 = 4

F2 ⊥ F3 donc minimum 4*4 = 16

F1 ⊥ F2, F1 ⊥ F3, F4 ⊥ F2 et F4 ⊥ F3 donc minimum 2*4 = 8

PPCM(4, 8, 16) = 16 essais

2. Quel modèle d'analyse peut-on utiliser ?

Pour analyser les résultats, on construit un modèle d'analyse de la variance à 4 facteurs :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl}$$

α_i= l'effet du lieu

β_j = l'effet du moment de la journée

γ_k = l'effet du type de fleur

δ_l = l'effet de la présence d'un arbre

ε_{ijkl} = la résiduelle

3. Construire un plan en 16 essais à la main ou à l'aide de la fonction `FrF2` du package `FrF2`. Saisir le plan construit sur Excel puis l'importer dans R. Quelles sont les propriétés de ce plan ?

Deux idées de résolution :

1) Construire un carré gréco-latin en juxtaposant deux carrés latins à 3 facteurs ; puis remplacer 2 variables à 4 modalités comme 2 variables à 2 modalités chacune en compressant deux niveaux.

2) Se servir de FrF2 : les facteurs à 4 modalités sont traités chacun comme 2 facteurs à 2 modalités.

```
plan <- FrF2(nruns=16, nfactors=6, factor.names=list(lieu=c("parc","jardin")
                  moment1=c("1","-1"), moment2=c("1","-1"),
                  plantes1=c("1","-1"), plantes2=c("1","-1"),
                  arbre=c("oui","non")))
```

```
     lieu  moment1 moment2 plantes1 plantes2 arbre
1     parc      -1       1       -1       -1   oui
2     parc       1      -1       -1       -1   non
3   jardin      -1      -1        1       -1   oui
4     parc       1       1        1        1   oui
5   jardin       1      -1       -1        1   oui
6     parc       1       1       -1        1   non
7     parc       1      -1        1       -1   oui
8     parc      -1      -1        1        1   non
9   jardin       1       1       -1       -1   oui
10    parc      -1      -1       -1        1   oui
11  jardin       1       1        1       -1   non
12  jardin       1      -1        1        1   non
13    parc      -1       1        1       -1   non
14  jardin      -1      -1       -1       -1   non
15  jardin      -1       1       -1        1   non
16  jardin      -1       1        1        1   oui
```

En combinant « à la main » deux niveaux comme suit :

```
 1   1    1
 1  -1    2
-1   1    3
-1  -1    4
```

On obtient le plan :

LIEU	MOMENT	PLANTE	ARBRE
PARC	15h	rose	oui
PARC	12h	rose	non

JARDIN	18h	eglantine	oui
PARC	9h	sauvage	oui
JARDIN	12h	marguerite	oui
PARC	9h	marguerite	non
PARC	12h	eglantine	oui
PARC	18h	sauvage	non
JARDIN	9h	rose	oui
PARC	18h	marguerite	oui
JARDIN	9h	eglantine	non
JARDIN	12h	sauvage	non
PARC	15h	eglantine	non
JARDIN	18h	rose	non
JARDIN	15h	marguerite	non
JARDIN	15h	sauvage	oui

On importe dans R. Puis on vérifie la qualité du plan :

```
options(contrasts=c("contr.sum","contr.sum"))
X1 <- model.matrix(~.,data=plan1)
round(solve(t(X1)%*%X1),2)

        (Intercept)lieu1 moment1 moment2 moment3 plante1 plante2 plante3 arbre1
(Intercept)    0.06  0.00    0.00    0.00    0.00    0.00    0.00    0.00   0.00
lieu1          0.00  0.06    0.00    0.00    0.00    0.00    0.00    0.00   0.00
moment1        0.00  0.00    0.19   -0.06   -0.06    0.00    0.00    0.00   0.00
moment2        0.00  0.00   -0.06    0.19   -0.06    0.00    0.00    0.00   0.00
moment3        0.00  0.00   -0.06   -0.06    0.19    0.00    0.00    0.00   0.00
plante1        0.00  0.00    0.00    0.00    0.00    0.19   -0.06   -0.06   0.00
plante2        0.00  0.00    0.00    0.00    0.00   -0.06    0.19   -0.06   0.00
plante3        0.00  0.00    0.00    0.00    0.00   -0.06   -0.06    0.19   0.00
arbre1         0.00  0.00    0.00    0.00    0.00    0.00    0.00    0.00   0.06
```

La matrice est presque diagonale. Certaines modalités des facteurs à 4 modalités sont légèrement confondues entre elles, les facteurs à 2 modalités ne sont pas liés aux autres facteurs.

Avec la fonction table, on voit que les lignes sont indépendantes. Comme les termes de la matrice ne sont pas constants, on parle de pseudo-orthogonalité.

```
table(plan1[,c(1,3)])
        plantes
lieu     eglantine marguerite rose sauvage
  jardin         2          2    2       2
  parc           2          2    2       2

table(plan1[,2:3])
      plantes
moment eglantine marguerite rose sauvage
   12h         1          1    1       1
   15h         1          1    1       1
   18h         1          1    1       1
   9h          1          1    1       1

table(plan1[,c(2,4)])
      arbre
moment non oui
   12h   2   2
   15h   2   2
   18h   2   2
   9h    2   2
```

4. **Utiliser la fonction `optFederov` du package `AlgDesign` pour construire un plan en 16 essais. Comparer ce plan avec celui construit à la main.**

On construit d'abord le plan complet :

```
pla <- fac.design(nlevels = (c(2,4,4,2)), randomize = FALSE,
factor.names=list(lieu=c("parc","jardin"),
```

```
                moment=c("9h","12h","15h","18h"),
                plant=c("sauvages","eglantine","marguerite","rose"),
                arbre=c("oui","non")))
```

On lance optFederov et on analyse la qualité du plan obtenu :

```
planopt <- optFederov(~.,data=pla,nTrials=16)
options(contrasts=c("contr.sum","contr.sum"))
Xo <- model.matrix(~.,planopt$design)
round(solve(t(Xo)%*%Xo),2)
```

```
        (Intercept)lieu1 moment1 moment2 moment3 plant1 plant2 plant3 arbre1
(Intercept)   0.06  0.00   0.00    0.00    0.00   0.00   0.00   0.00   0.00
lieu1         0.00  0.06   0.00    0.00    0.00   0.00   0.00   0.00   0.00
moment1       0.00  0.00   0.19   -0.06   -0.06   0.00   0.00   0.00   0.00
moment2       0.00  0.00  -0.06    0.19   -0.06   0.00   0.00   0.00   0.00
moment3       0.00  0.00  -0.06   -0.06    0.19   0.00   0.00   0.00   0.00
plant1        0.00  0.00   0.00    0.00    0.00   0.19  -0.06  -0.06   0.00
plant2        0.00  0.00   0.00    0.00    0.00  -0.06   0.19  -0.06   0.00
plant3        0.00  0.00   0.00    0.00    0.00  -0.06  -0.06   0.19   0.00
arbre1        0.00  0.00   0.00    0.00    0.00   0.00   0.00   0.00   0.06
```

On obtient la même matrice que précédemment : les plans sont équivalents.

Remarque : on obtient aussi cette même matrice avec `oa.design`.

<u>PARTIE 2</u> : Construction d'un plan continu

On se rend compte que certaines modalités choisies pour les précédentes expériences dans le jardin ne sont pas forcément optimales. Par exemple, on avait choisi au hasard une zone de la pelouse et on se demande si les caractéristiques du quadrat influence l'abondance des insectes.

On s'intéresse désormais aux facteurs quantitatifs suivants :

- La localisation du quadrat dans le jardin (1) : entre 5 et 20 m de la clôture Sud
- La localisation du quadrat dans le jardin (2) : entre 5 et 20 m de la clôture Ouest
- Le pourcentage de trèfle dans le quadrat : entre 20 et 70 %

On cherche pour quelles modalités le nombre d'insectes est maximal.

5. **Proposer un plan d'expérience à l'aide du package rsm.**

```
plan <- ccd(3, randomize=FALSE, coding=list (x1~(Sud-12.5)/7.5,  x2~(Ouest-12.5)/7.5),
x3~(trefle-45)/25, alpha="rotatable")
```

On obtient le plan suivant à 22 essais avec 8 points au centre :

```
   run.order std.order        Sud      Ouest    trefle Block
1          1         1  5.0000000  5.0000000 20.000000     1
2          2         2 20.0000000  5.0000000 20.000000     1
3          3         3  5.0000000 20.0000000 20.000000     1
4          4         4 20.0000000 20.0000000 20.000000     1
5          5         5  5.0000000  5.0000000 70.000000     1
6          6         6 20.0000000  5.0000000 70.000000     1
7          7         7  5.0000000 20.0000000 70.000000     1
8          8         8 20.0000000 20.0000000 70.000000     1
9          9         9 12.5000000 12.5000000 45.000000     1
10        10        10 12.5000000 12.5000000 45.000000     1
11        11        11 12.5000000 12.5000000 45.000000     1
12        12        12 12.5000000 12.5000000 45.000000     1
13         1         1 -0.1134462 12.5000000 45.000000     2
14         2         2 25.1134462 12.5000000 45.000000     2
15         3         3 12.5000000 -0.1134462 45.000000     2
16         4         4 12.5000000 25.1134462 45.000000     2
17         5         5 12.5000000 12.5000000  2.955179     2
18         6         6 12.5000000 12.5000000 87.044821     2
19         7         7 12.5000000 12.5000000 45.000000     2
20         8         8 12.5000000 12.5000000 45.000000     2
21         9         9 12.5000000 12.5000000 45.000000     2
22        10        10 12.5000000 12.5000000 45.000000     2
```

6. **Les données récoltées sont stockées dans le vecteur suivant : nombre d'insectes**

   Y <- c(7,11,11,13,19,16,18,20,18,19,15,15,16,17,15,22,5,25,17,17,20,19)

   Analyser les résultats. Quel modèle doit-on conserver ?

```
rep <- rsm(Y~SO(x1,x2,x3),data=plan)
summary(rep)

Call:
rsm(formula = Y ~ SO(x1, x2, x3), data = plan)

            Estimate Std. Error t value  Pr(>|t|)    
(Intercept) 17.55430    0.81444 21.5538 5.813e-11 ***
x1           0.48926    0.62380  0.7843   0.44806    
x2           1.52103    0.62380  2.4383   0.03125 *  
x3           4.73285    0.62380  7.5871 6.441e-06 ***
x1:x2        0.37500    0.81504  0.4601   0.65367    
x1:x3       -0.87500    0.81504 -1.0736   0.30413    
x2:x3       -0.37500    0.81504 -0.4601   0.65367    
x1^2        -0.82032    0.58569 -1.4006   0.18666    
x2^2        -0.11321    0.58569 -0.1933   0.84997    
x3^2        -1.35065    0.58569 -2.3061   0.03976 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.8585,	Adjusted R-squared:  0.7523 
F-statistic: 8.088 on 9 and 12 DF,  p-value: 0.0006928

Analysis of Variance Table

Response: Y
                Df Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2, x3)   3 340.78 113.592 21.3747 4.185e-05 #effets linéaires
TWI(x1, x2, x3)  3   8.38   2.792  0.5253    0.6731 #effets d’interactions
PQ(x1, x2, x3)   3  37.67  12.556  2.3626    0.1225 #effets quadratiques
Residuals       12  63.77   5.314                  
Lack of fit      5  39.77   7.954  2.3200    0.1514 #erreur d’ajustement
Pure error       7  24.00   3.429                  
```

On commence par étudier le tableau d’analyse de la variance des effets globaux.

La p-value de l’erreur d’ajustement est supérieure à 0,05 donc le modèle est bien ajusté (ajustement global R^2 = 0,75 ce qui est très satisfaisant). Les effets linéaires sont très significatifs. On voit que l’effet du trèfle est prépondérant sur le nombre d’insectes observés. L’éloignement de la clôture Ouest est également significatif, mais aussi son effet quadratique.

Ainsi on peut conserver un modèle sans effets d’interactions.

7. Construire le nouveau modèle. Conclure.

```
ajust <- rsm(Y~FO(x1,x2,x3)+PQ(x1,x2,x3),data=plan)
summary(ajust)

Call:
rsm(formula = Y ~ FO(x1, x2, x3) + PQ(x1, x2, x3), data = plan)

            Estimate Std. Error t value  Pr(>|t|)    
(Intercept) 17.55430    0.77482 22.6561 5.144e-13 ***
x1           0.48926    0.59346  0.8244   0.42261    
x2           1.52103    0.59346  2.5630   0.02163 *  
x3           4.73285    0.59346  7.9751 8.948e-07 ***
x1^2        -0.82032    0.55720 -1.4722   0.16163    
x2^2        -0.11321    0.55720 -0.2032   0.84173    
x3^2        -1.35065    0.55720 -2.4240   0.02846 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.8399,	Adjusted R-squared:  0.7758 
F-statistic: 13.11 on 6 and 15 DF,  p-value: 3.215e-05

Analysis of Variance Table

Response: Y
               Df Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2, x3)  3 340.78 113.592 23.6169 6.195e-06
PQ(x1, x2, x3)  3  37.67  12.556  2.6105   0.08973
Residuals      15  72.15   4.810                  
Lack of fit     8  48.15   6.018  1.7554   0.23638
Pure error      7  24.00   3.429                  

Stationary point of response surface:
       x1        x2        x3 
0.2982164 6.7178383 1.7520700 

Stationary point in original units:
```

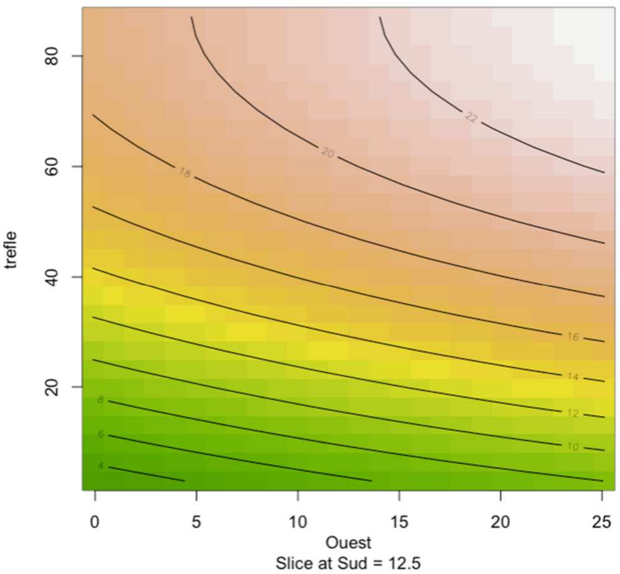
```
      Sud     Ouest    trefle
14.73662 62.88379 88.80175

Eigenanalysis:
eigen() decomposition
$values
[1] -0.1132086 -0.8203154 -1.3506455

$vectors
   [,1] [,2] [,3]
x1    0    1    0
x2    1    0    0
x3    0    0    1
```

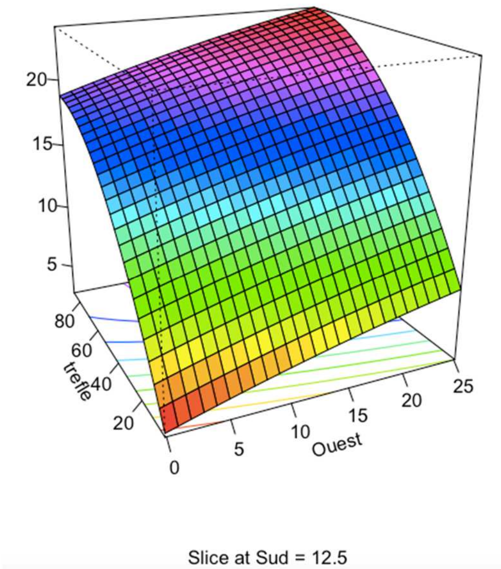
Ainsi on voit que le modèle est encore mieux ajusté (ajustement global R^2 = 0,78 ce qui est très satisfaisant).

Les valeurs propres sont toutes négatives, l'optimum est donc un maximum. Ainsi pour maximiser l'observation d'insectes, il faut étudier des parcelles de trèfles loin de la clôture Ouest.

Ce qui est confirmé par :

```
contour(ajust,~x1+x2+x3,image=TRUE)
persp(ajust,~x1+x2+x3,contours="colors")
```





Exercice 4. Revêtement d'intérieur de piscine (exemple fictif, par Emma Mercier)

Partie 1 : Plan fractionnaire

Thierry vend des peintures qui servent de revêtement pour l'intérieur des bassins de piscine en béton. Solution privilégiée par les particuliers pour revêtir leur piscine, son affaire tourne mais les retours après-vente sont parfois d'avis différents : la peinture cloque chez certains mais quasiment pas chez d'autres.

Il sait que sa peinture de piscine doit principalement supporter l'immersion dans l'eau : ceci est permis par la présence d'adjuvant hydrofuge en poudre ou liquide. De plus, la peinture de piscine peut être composé soit de caoutchouc soit de résine de synthèse. Ces deux composés permettent de faciliter son étalement. Ensuite, cette peinture doit être applicable sur un béton neuf ou ancien. L'intérieur de la piscine doit cependant être préalablement lavé par de l'acide chlorhydrique ou de l'eau de javel diluée. Thierry s'intéresse donc à l'application de sa peinture selon 4 facteurs à 2 modalités :

- F1 : La composition liée à l'étalement : en caoutchouc (1) ou en résine de synthèse (2)
- F2 : La composition liée à l'étanchéité : ajout d'adjuvant hydrofuge en poudre (1) ou liquide (2)
- F3 : La surface d'application : béton neuf (1) ou béton ancien (2)
- F4 : Lavage de la surface à peindre : acide chlorhydrique (1) ou eau de javel diluée (2)

Ainsi, il a essayé de comprendre les difficultés que rencontrent les particuliers avec sa peinture en se limitant à 8 essais sur une surface de même taille pour chaque essai. Le but de l'exercice est donc d'aider Thierry dans sa démarche. Thierry ne connaît aucun outil informatique et propose le design d'expériences suivant :

Essai	Étalement	Étanchéité	Béton	Lavage

1	1	1	1	1
2	1	1	2	2
3	1	2	1	2
4	1	2	2	1
5	2	1	1	2
6	2	1	2	1
7	2	2	1	1
8	2	2	2	2

1. Indiquer la variable à expliquer dans la démarche de Thierry. Écrire le modèle linéaire associé, c'est-à-dire sans prendre en compte les interactions, et le plan fractionnaire à construire.

La variable à expliquer est le pourcentage de surface de peinture qui cloque sur la surface totale à recouvrir de peinture. Une analyse de variance à 4 facteurs permettrait d'analyser les résultats. Le modèle linéaire s'écrit :

$$\forall i,j,k,l \qquad Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl}$$

avec $\mathcal{L}(\varepsilon_{ijkl}) = \mathcal{N}(0,\sigma)$ et $\mathrm{Cov}(\varepsilon_{ijkl}, \varepsilon_{i'j'k'l'}) = 0 \quad \forall (i',j',k',l') \neq (i,j,k,l)$.

Thierry souhaite réaliser 8 essais avec 4 facteurs à 2 modalités. Le plan fractionnaire à construire est le plan 2^{4-1}.

2. Proposez une matrice des essais du plan fractionnaire à partir du design de l'expérience de Thierry. Déterminez le générateur d'alias et la résolution du plan. Commentez.

En remplaçant la valeur « 2 » par « -1 » de chaque modalité de chaque facteur, on obtient une matrice des essais suivante :

F1	F2	F3	F4
1	1	1	1
1	1	-1	-1
1	-1	1	-1
1	-1	-1	1
-1	1	1	-1
-1	1	-1	1
-1	-1	1	1
-1	-1	-1	-1

On remarque que les trois premiers facteurs forment un plan complet. Comme on s'intéresse également au 4ème facteur : F4, on le confond avec l'interaction d'ordre 3, F1F2F3, qu'on supposera négligeable. On suppose donc que le facteur F4 et l'interaction F1F2F3 ne pourront pas être distingués. Dans ce cas, le seul générateur d'alias est I= F1F2F3F4. Le plan fractionnaire est donc de résolution IV.

Si on utilise le générateur d'alias, on voit que tous les effets principaux F1, F2 et F3 sont confondus avec des interactions d'ordre 3 supposées négligeables. Ainsi, si les résultats obtenus par Thierry, c'est-à-dire les pourcentages de surface qui cloquent, varient au cours des essais cela est dû aux effets principaux (qui bougent) et non à cause des effets d'interaction d'ordre 3 (qui bougent aussi au cours des expériences). De plus, les interactions d'ordre 2 sont également confondues entre elles donc en 8 essais, en résolution IV, Thierry peut estimer correctement les 4 effets principaux (F1, F2, F3 et F4).

3. Construire le plan sous R selon le design de Thierry en utilisant la fonction *FrF2*. Calculer la matrice $(X'X)^{-1}$ quand vous utilisez ce plan et que vous construisez un modèle avec seulement les effets principaux. Commentez et proposez une modification du design de Thierry pour en améliorer la qualité.

```
library(FactoMineR)
library(FrF2)
F1 <- as.factor(c(1,1,-1,-1,-1,-1,1,1))
F2 <- as.factor(c(-1,-1,-1,1,-1,1,1,1))
F3 <- as.factor(c(-1,1,1,-1,-1,1,-1,1))
F4 <- as.factor(c(1,-1,1,1,-1,-1,-1,1))
don <- cbind.data.frame(F1,F2,F3,F4)
X <- model.matrix(~., data = don)
t(X)%*%X
             (Intercept) F11 F21 F31 F41
(Intercept)            8   4   4   4   4
F11                    4   4   2   2   2
F21                    4   2   4   2   2
F31                    4   2   2   4   2
F41                    4   2   2   2   4
```

```
solve(t(X)%*%X)
            (Intercept)   F11   F21   F31   F41
(Intercept)       0.625 -0.25 -0.25 -0.25 -0.25
F11              -0.250  0.50  0.00  0.00  0.00
F21              -0.250  0.00  0.50  0.00  0.00
F31              -0.250  0.00  0.00  0.50  0.00
F41              -0.250  0.00  0.00  0.00  0.50
```

Pour avoir un plan fractionnaire de qualité avec un nombre d'essais donnés, il faut chercher des expériences qui minimisent $(X'X)^{-1}$. Or, on sait que dans un plan fractionnaire, les facteurs sont orthogonaux. Ainsi, la matrice $(X'X)^{-1}$ est diagonale et dans le cas de Thierry, la valeur de la diagonale peut atteindre le minimum de 1/8 (8 essais) soit 0,125. Selon le design de Thierry, on peut voir qu'il y a une augmentation de la valeur de la variance de l'estimateur (0,50>0,125). Il n'a donc pas choisi les « bonnes » 8 expériences parmi les 16 possibles du plan complet 2^4. On peut tout de même proposer le design suivant pour effectuer les 8 essais qui permettent de minimiser la valeur de l'estimateur. Thierry pourrait envisager d'augmenter le nombre d'essais pour mieux estimer les effets principaux et ceux d'interactions d'ordre supérieur à 2.

```
plan <- FrF2(nfactors = 4, nruns = 8)
summary(plan)

Call:
FrF2(nfactors = 4, nruns = 8)

Experimental design of type  FrF2
8  runs

Factor settings (scale ends):
   A  B  C  D
1 -1 -1 -1 -1
2  1  1  1  1

Design generating information:
$legend
[1] A=A B=B C=C D=D

$generators
[1] D=ABC

Alias structure:
$fi2
[1] AB=CD AC=BD AD=BC

The design itself:
   A  B  C  D
1 -1  1 -1  1
2  1 -1 -1  1
3 -1  1  1 -1
4  1 -1  1 -1
5 -1 -1  1  1
6 -1 -1 -1 -1
7  1  1 -1 -1
8  1  1  1  1
class=design, type= FrF2

X <- model.matrix(~., data = plan)
t(X)%*%X
solve(t(X)%*%X)

            (Intercept) A1 B1 C1 D1
(Intercept)           8  0  0  0  0
A1                    0  8  0  0  0
B1                    0  0  8  0  0
C1                    0  0  0  8  0
D1                    0  0  0  0  8


            (Intercept)    A1    B1    C1    D1
(Intercept)       0.125 0.000 0.000 0.000 0.000
A1                0.000 0.125 0.000 0.000 0.000
B1                0.000 0.000 0.125 0.000 0.000
C1                0.000 0.000 0.000 0.125 0.000
D1                0.000 0.000 0.000 0.000 0.125
```

4. Proposez à Thierry un nouveau plan pour mieux estimer les effets principaux et les effets d'interactions éventuels.

S'il passe à 16 essais, on peut construire un plan fractionnaire 2^4, qui correspond au plan complet du modèle et permettra d'estimer les effets principaux et les effets d'interactions.

5. **Bonne nouvelle ! Thierry a suivi vos conseils. En utilisant R, calculez la nouvelle matrice $(X'X)^{-1}$ de ce nouveau plan en mettant en avant les interactions d'ordre 2 puis commentez.**

```
plan <- FrF2(nfactors = 4, nruns = 16)
summary(plan)

Call:
FrF2(nfactors = 4, nruns = 16)

Experimental design of type  FrF2
16  runs

Factor settings (scale ends):
   A  B  C  D
1 -1 -1 -1 -1
2  1  1  1  1

The design itself:
    A  B  C  D
1   1  1 -1 -1
2   1  1 -1  1
3   1 -1  1 -1
4  -1  1  1 -1
5   1 -1 -1 -1
6  -1  1  1  1
7  -1 -1 -1  1
8   1  1  1  1
9   1 -1 -1  1
10  1 -1  1  1
11 -1  1 -1  1
12  1  1  1 -1
13 -1 -1 -1 -1
14 -1  1 -1 -1
15 -1 -1  1 -1
16 -1 -1  1  1
class=design, type= FrF2
X <- model.matrix(~.+A:B+A:C+A:D+B:C+B:D+C:D, data = plan)
t(X)%*%X
solve(t(X)%*%X)

            (Intercept) A1 B1 C1 D1 A1:B1 A1:C1 A1:D1 B1:C1 B1:D1 C1:D1
(Intercept)          16  0  0  0  0     0     0     0     0     0     0
A1                    0 16  0  0  0     0     0     0     0     0     0
B1                    0  0 16  0  0     0     0     0     0     0     0
C1                    0  0  0 16  0     0     0     0     0     0     0
D1                    0  0  0  0 16     0     0     0     0     0     0
A1:B1                 0  0  0  0  0    16     0     0     0     0     0
A1:C1                 0  0  0  0  0     0    16     0     0     0     0
A1:D1                 0  0  0  0  0     0     0    16     0     0     0
B1:C1                 0  0  0  0  0     0     0     0    16     0     0
B1:D1                 0  0  0  0  0     0     0     0     0    16     0
C1:D1                 0  0  0  0  0     0     0     0     0     0    16

            (Intercept)     A1     B1     C1     D1  A1:B1  A1:C1  A1:D1  B1:C1  B1:D1  C1:D1
(Intercept)      0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
A1               0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
B1               0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
C1               0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
D1               0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
A1:B1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000
A1:C1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000
A1:D1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000
B1:C1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000
B1:D1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000
C1:D1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625
```

Il réalise donc un plan fractionnaire avec 16 essais et 4 facteurs. Ainsi, la nouvelle matrice $(X'X)^{-1}$ est toujours diagonale, même en tenant compte des effets d'interactions d'ordre 2 entre les 4 premiers facteurs. A ce stade, il n'y a donc plus de confusions d'effets associés à ce plan qui est complet. Thierry pourra alors mieux estimer les coefficients du modèle sans confusion d'effets.

6. **En suivant vos conseils, Thierry a donc obtenu les pourcentages de surface qui cloquent suivants de l'essai 1 à 16 : 8,20,3,6,15,2,1,3,12,35,13,6,3,15,3,23. A l'aide de R, aidez-le à interpréter ses résultats en faisant apparaître les interactions d'ordre 2 et 3.**

```
plan <- FrF2(nfactors = 4, nruns = 16)
summary(plan)

Call:
FrF2(nfactors = 4, nruns = 16)
```

```
Experimental design of type  FrF2
16  runs

Factor settings (scale ends):
   A  B  C  D
1 -1 -1 -1 -1
2  1  1  1  1

The design itself:
    A  B  C  D
1  -1 -1  1 -1
2   1 -1  1 -1
3   1 -1 -1 -1
4  -1 -1 -1  1
5  -1 -1  1  1
6  -1  1 -1 -1
7   1  1  1 -1
8  -1  1 -1  1
9   1  1  1  1
10  1 -1 -1  1
11 -1  1  1 -1
12 -1 -1 -1 -1
13  1  1 -1 -1
14  1 -1  1  1
15 -1  1  1  1
16  1  1 -1  1
class=design, type= full factorial
Y <- c(8,20,3,6,15,2,1,3,12,35,13,6,3,15,3,23)
AovSum(Y~.+A:B+A:C+A:D+B:C+B:D+C:D=A:B:C+A:B:D+A:C:D+B:C:D, data= plan)
Ftest
              SS df     MS F value  Pr(>F)
A           4.00  1   4.00    4.00 0.29517
B         324.00  1 324.00  324.00 0.03533 *
C           6.25  1   6.25    6.25 0.24224
D          12.25  1  12.25   12.25 0.17717
A:B         4.00  1   4.00    4.00 0.29517
A:C        72.25  1  72.25   72.25 0.07455 .
A:D         2.25  1   2.25    2.25 0.37433
B:C         2.25  1   2.25    2.25 0.37433
B:D        20.25  1  20.25   20.25 0.13921
C:D       256.00  1 256.00  256.00 0.03974 *
A:B:C     110.25  1 110.25  110.25 0.06045 .
A:B:D      30.25  1  30.25   30.25 0.11450
A:C:D     196.00  1 196.00  196.00 0.04540 *
B:C:D     289.00  1 289.00  289.00 0.03741 *
Residuals   1.00  1   1.00
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ttest
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       10.500      0.250    42.0  0.01515 *
A - -1            -0.500      0.250    -2.0  0.29517
A - 1              0.500      0.250     2.0  0.29517
B - -1             4.500      0.250    18.0  0.03533 *
B - 1             -4.500      0.250   -18.0  0.03533 *
C - -1             0.625      0.250     2.5  0.24224
C - 1             -0.625      0.250    -2.5  0.24224
D - -1             0.875      0.250     3.5  0.17717
D - 1             -0.875      0.250    -3.5  0.17717
A - -1 : B - -1   -0.500      0.250    -2.0  0.29517
A - 1 : B - -1     0.500      0.250     2.0  0.29517
A - -1 : B - 1     0.500      0.250     2.0  0.29517
A - 1 : B - 1     -0.500      0.250    -2.0  0.29517
A - -1 : C - -1    2.125      0.250     8.5  0.07455 .
A - 1 : C - -1    -2.125      0.250    -8.5  0.07455 .
A - -1 : C - 1    -2.125      0.250    -8.5  0.07455 .
A - 1 : C - 1      2.125      0.250     8.5  0.07455 .
A - -1 : D - -1    0.375      0.250     1.5  0.37433
A - 1 : D - -1    -0.375      0.250    -1.5  0.37433
A - -1 : D - 1    -0.375      0.250    -1.5  0.37433
A - 1 : D - 1      0.375      0.250     1.5  0.37433
B - -1 : C - -1   -0.375      0.250    -1.5  0.37433
B - 1 : C - -1     0.375      0.250     1.5  0.37433
B - -1 : C - 1     0.375      0.250     1.5  0.37433
B - 1 : C - 1     -0.375      0.250    -1.5  0.37433
B - -1 : D - -1    1.125      0.250     4.5  0.13921
B - 1 : D - -1    -1.125      0.250    -4.5  0.13921
B - -1 : D - 1    -1.125      0.250    -4.5  0.13921
B - 1 : D - 1      1.125      0.250     4.5  0.13921
C - -1 : D - -1   -4.000      0.250   -16.0  0.03974 *
C - 1 : D - -1     4.000      0.250    16.0  0.03974 *
C - -1 : D - 1     4.000      0.250    16.0  0.03974 *
C - 1 : D - 1     -4.000      0.250   -16.0  0.03974 *
```

```
A - -1 : B - -1    2.625      0.250    10.5  0.06045 .
A - 1 : B - -1    -2.625      0.250   -10.5  0.06045 .
A - -1 : B - 1    -2.625      0.250   -10.5  0.06045 .
A - 1 : B - 1      2.625      0.250    10.5  0.06045 .
A - -1 : B - -1   -1.375      0.250    -5.5  0.11450
A - 1 : B - -1     1.375      0.250     5.5  0.11450
A - -1 : B - 1     1.375      0.250     5.5  0.11450
A - 1 : B - 1     -1.375      0.250    -5.5  0.11450
A - -1 : C - -1    3.500      0.250    14.0  0.04540 *
A - 1 : C - -1    -3.500      0.250   -14.0  0.04540 *
A - -1 : C - 1    -3.500      0.250   -14.0  0.04540 *
A - 1 : C - 1      3.500      0.250    14.0  0.04540 *
B - -1 : C - -1   -4.250      0.250   -17.0  0.03741 *
B - 1 : C - -1     4.250      0.250    17.0  0.03741 *
B - -1 : C - 1     4.250      0.250    17.0  0.03741 *
B - 1 : C - 1     -4.250      0.250   -17.0  0.03741 *
---
Signif.  codes:    0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1  >  X  <-
model.matrix(Y~.+A:B+A:C+A:D+B:C+B:D+C:D A:B:C+A:B:D+A:C:D+B:C:D, data = plan)
X <- model.matrix(Y~.+A:B+A:C+A:D+B:C+B:D+C:D+A:B:C+A:B:D+A:C:D+B:C:D, data = plan)
t(X)%*%X
solve(t(X)%*%X)summary(plan)
            (Intercept) A1 B1 C1 D1 A1:B1 A1:C1 A1:D1 B1:C1 B1:D1 C1:D1 A1:B1:C1 A1:B1:D1 A1:C1:D1 B1:C1:D1
(Intercept)          16  0  0  0  0     0     0     0     0     0     0        0        0        0        0
A1                    0 16  0  0  0     0     0     0     0     0     0        0        0        0        0
B1                    0  0 16  0  0     0     0     0     0     0     0        0        0        0        0
C1                    0  0  0 16  0     0     0     0     0     0     0        0        0        0        0
D1                    0  0  0  0 16     0     0     0     0     0     0        0        0        0        0
A1:B1                 0  0  0  0  0    16     0     0     0     0     0        0        0        0        0
A1:C1                 0  0  0  0  0     0    16     0     0     0     0        0        0        0        0
A1:D1                 0  0  0  0  0     0     0    16     0     0     0        0        0        0        0
B1:C1                 0  0  0  0  0     0     0     0    16     0     0        0        0        0        0
B1:D1                 0  0  0  0  0     0     0     0     0    16     0        0        0        0        0
C1:D1                 0  0  0  0  0     0     0     0     0     0    16        0        0        0        0
A1:B1:C1              0  0  0  0  0     0     0     0     0     0     0       16        0        0        0
A1:B1:D1              0  0  0  0  0     0     0     0     0     0     0        0       16        0        0
A1:C1:D1              0  0  0  0  0     0     0     0     0     0     0        0        0       16        0
B1:C1:D1              0  0  0  0  0     0     0     0     0     0     0        0        0        0       16

            (Intercept)     A1     B1     C1     D1  A1:B1  A1:C1  A1:D1  B1:C1  B1:D1  C1:D1 A1:B1:C1 A1:B1:D1 A1:C1:D1 B1:C1:D1
(Intercept)      0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   0.0000   0.0000   0.0000   0.0000
A1               0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   0.0000   0.0000   0.0000   0.0000
B1               0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   0.0000   0.0000   0.0000   0.0000
C1               0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   0.0000   0.0000   0.0000   0.0000
D1               0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   0.0000   0.0000   0.0000   0.0000
A1:B1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000   0.0000   0.0000   0.0000   0.0000
A1:C1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000   0.0000   0.0000   0.0000   0.0000
A1:D1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000   0.0000   0.0000   0.0000   0.0000
B1:C1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000   0.0000   0.0000   0.0000   0.0000
B1:D1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000   0.0000   0.0000   0.0000   0.0000
C1:D1            0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625   0.0000   0.0000   0.0000   0.0000
A1:B1:C1         0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   0.0625   0.0000   0.0000   0.0000
A1:B1:D1         0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   0.0000   0.0625   0.0000   0.0000
A1:C1:D1         0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   0.0000   0.0000   0.0625   0.0000
B1:C1:D1         0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000   0.0000   0.0000   0.0000   0.0625
```

La fonction *AovSum* permet de déterminer quels facteurs ont un effet non négligeable avec une probabilité critique inférieure à 0,05 (*) sur la variable Y et la valeur des coefficients qui constituent le modèle. On comprend alors que la constante, les facteurs F2 (B), les interactions F3F4 (CD), F1F3F4 (ACD) et F2F3F4 (BCD) ont des coefficients significativement différents de 0. Ces éléments sont donc les facteurs qui expliquent le plus la variabilité de Y(pourcentage de surface qui cloque) du modèle prenant en compte l'ensemble des interactions. Thierry devrait donc se concentrer plus particulièrement sur ces 5 éléments pour comprendre pourquoi la peinture cloque.

Partie 2 : Plan asymétrique

Thierry s'est renseigné auprès de ses clients et de son équipe. Il apprend alors que les bassins de piscine peuvent également être en ciment et lavé seulement avec de l'eau mais également que le temps qui sépare l'application et le remplissage a varié selon les clients. Certains ont donc rempli leur piscine après 24h, 36h, 48h et 72h après l'application de la peinture.

Thierry s'intéresse donc à l'application de sa peinture selon 5 facteurs à 2, 3 et 4 modalités :

- F1 : La composition liée à l'étalement : en caoutchouc (1) ou en résine de synthèse (2)
- F2 : La composition liée à l'étanchéité : ajout d'adjuvant hydrofuge en poudre (1) ou liquide (2)
- F3 : La surface d'application : béton neuf (1) ou béton ancien (2) ou en ciment (3)
- F4 : Lavage de la surface à peindre : acide chlorhydrique (1) ou eau de javel diluée (2) ou de l'eau (3)
- F5 : le temps qui sépare l'application et le remplissage : 24h (1), 36h (2), 48h (3) et 72h (4).

7. Déterminer le nombre d'essais maximum et minimum à réaliser dans ce cas. Commentez.

On a 2 facteurs à 2 modalités, 2 facteurs à 3 modalités et 1 facteur à 4 modalités. Il y a donc 2*2*3*3*4 = 144 essais maximum à réaliser. Le nombre minimum d'essais à réaliser est lié au degré de liberté apporté par chaque facteur : 1 (constante) + 2 *

(2-1) + 2 * (3-1) + (4-1) = 10 essais minimums. Pour avoir l'orthogonalité, il faut que chaque facteur soit orthogonal 2 à 2. Il faut donc déterminer le PPCM (4 ; 6 ; 8 ; 9 ; 12) = 72. Ainsi, 72 essais permettraient d'avoir un plan orthogonal (sans confusions d'effets)

8. **En vous ramenant aux plans fractionnaires classiques (méthode 1) et en compressant le facteur à 4 niveaux en 3 (méthode 2), définissez le nom du plan ($L_a2^b3^c$) où a correspond au nombre d'essais pour estimer sans confusion d'effets (orthogonalité) le modèle à 5 facteurs à 2,3 et 4 modalités pour chaque méthode. Vous préciserez les plans fractionnaires « intermédiaires » à utiliser pour construire la planification expérimentale pour chaque méthode.**

Méthode 1 : On peut construire un facteur à 4 modalités à partir de 2 facteurs à 2 modalités. On se ramène à étudier 4 facteurs à 2 modalités et 2 facteurs à 3 modalités. Ensuite, pour construire la planification expérimentale sur 36 essais, on construira deux plans fractionnaires : le plan 2^{4-2} pour les 4 facteurs à 2 modalités et le plan complet 3^2 pour les 2 facteurs à 3 modalités. Ces deux plans forment donc le plan $L_{36}2^43^2$.

Méthode 2 : On peut compresser des facteurs à 4 modalités en facteurs à 3 modalités. On se ramène à étudier 2 facteurs à 2 modalités et 2 facteurs à 4 modalités, soit 8 facteurs à 2 modalités. On utilisera alors 6 de ces facteurs pour construire 3 facteurs à 4 modalités, puis on utiliser 2 de ces facteurs à 4 modalités pour compresser les niveaux et obtenir 2 facteurs à 3 modalités. On construira alors un plan fractionnaires 2^{8-4}.

9. **A l'aide du critère D-optimalité, déterminer quel plan optimal $L_a2^43^2$ est à préconiser (avec un nombre d'essais égal à 36,72 puis 144).**

Pour Thierry, il est préférable de réaliser un nombre d'expériences limité (36 avec les méthodes expliquées précédemment car 144 expériences c'est couteux et fastidieux). De plus, pour que le plan soit de qualité, il faut essayer d'avoir une matrice $(X'X)^{-1}$ diagonale pour limiter les confusions d'effets. On peut comparer la valeur du déterminant de la matrice pour chaque nombre d'essais.

Méthode 1 le plan $L_{36}2^43^2$:

```
library(FactoMineR)
library(FrF2)
library(DoE.base)
plan2 <- oa.design(nlevels = c(2,2,2,2,3,3), nruns=36)
X <- model.matrix(~., data = plan2)
t(X)%*%X
            (Intercept) A1 B1 C1 D1 E.L E.Q F.L F.Q
(Intercept)          36  0  0  0  0   0   0   0   0
A1                    0 36  0  0  0   0   0   0   0
B1                    0  0 36  0  0   0   0   0   0
C1                    0  0  0 36  0   0   0   0   0
D1                    0  0  0  0 36   0   0   0   0
E.L                   0  0  0  0  0  12   0   0   0
E.Q                   0  0  0  0  0   0  12   0   0
F.L                   0  0  0  0  0   0   0  12   0
F.Q                   0  0  0  0  0   0   0   0  12
round(solve(t(X)%*%X),3)
            (Intercept)    A1    B1    C1    D1   E.L   E.Q   F.L   F.Q
(Intercept)       0.028 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
A1                0.000 0.028 0.000 0.000 0.000 0.000 0.000 0.000 0.000
B1                0.000 0.000 0.028 0.000 0.000 0.000 0.000 0.000 0.000
C1                0.000 0.000 0.000 0.028 0.000 0.000 0.000 0.000 0.000
D1                0.000 0.000 0.000 0.000 0.028 0.000 0.000 0.000 0.000
E.L               0.000 0.000 0.000 0.000 0.000 0.083 0.000 0.000 0.000
E.Q               0.000 0.000 0.000 0.000 0.000 0.000 0.083 0.000 0.000
F.L               0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.083 0.000
F.Q               0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.083

 round(det(solve(t(X)%*%X)),3)
 [1] 7.976e-13
```

La matrice de dispersion $(X'X)^{-1}$ de la méthode 1 a un déterminant plus petit que celui de la matrice de la méthode 2. Donc le plan $L_{36}2^43^2$ est le plan optimal pour construire une planification expérimentale en 36 essais en se ramenant à 4 facteurs à 2 modalités et 2 facteurs à 3 modalités.

le plan $L_{72}2^43^2$:

```
library(FactoMineR)
library(FrF2)
library(DoE.base)
plan2 <- oa.design(nlevels = c(2,2,2,2,3,3), nruns=72)
X <- model.matrix(~., data = plan2)
round(t(X)%*%X,2)
            (Intercept) A1 B1 C1 D1 E.L E.Q F.L F.Q
(Intercept)          72  0  0  0  0   0   0   0   0
A1                    0 72  0  0  0   0   0   0   0
B1                    0  0 72  0  0   0   0   0   0
```

```
C1                  0   0   0  72   0    0    0    0    0
D1                  0   0   0   0  72    0    0    0    0
E.L                 0   0   0   0   0   24    0    0    0
E.Q                 0   0   0   0   0    0   24    0    0
F.L                 0   0   0   0   0    0    0   24    0
F.Q                 0   0   0   0   0    0    0    0   24
solve(t(X)%*%X)
            (Intercept)    A1    B1    C1    D1   E.L   E.Q   F.L   F.Q
(Intercept)       0.014 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
A1                0.000 0.014 0.000 0.000 0.000 0.000 0.000 0.000 0.000
B1                0.000 0.000 0.014 0.000 0.000 0.000 0.000 0.000 0.000
C1                0.000 0.000 0.000 0.014 0.000 0.000 0.000 0.000 0.000
D1                0.000 0.000 0.000 0.000 0.014 0.000 0.000 0.000 0.000
E.L               0.000 0.000 0.000 0.000 0.000 0.042 0.000 0.000 0.000
E.Q               0.000 0.000 0.000 0.000 0.000 0.000 0.042 0.000 0.000
F.L               0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.042 0.000
F.Q               0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.042
round(det(solve(t(X)%*%X)),16)
 [1] 1.6e-15
```

le plan $L_{144}2^4 3^2$:

```
library(FactoMineR)
library(FrF2)
library(DoE.base)
plan2 <- oa.design(nlevels = c(2,2,2,2,3,3), nruns=144)
X <- model.matrix(~., data = plan2)
t(X)%*%X
            (Intercept)  A1  B1  C1  D1 E.L E.Q F.L F.Q
(Intercept)         144   0   0   0   0   0   0   0   0
A1                    0 144   0   0   0   0   0   0   0
B1                    0   0 144   0   0   0   0   0   0
C1                    0   0   0 144   0   0   0   0   0
D1                    0   0   0   0 144   0   0   0   0
E.L                   0   0   0   0   0  48   0   0   0
E.Q                   0   0   0   0   0   0  48   0   0
F.L                   0   0   0   0   0   0   0  48   0
F.Q                   0   0   0   0   0   0   0   0  48
solve(t(X)%*%X)
            (Intercept)    A1    B1    C1    D1   E.L   E.Q   F.L   F.Q
(Intercept)       0.007 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
A1                0.000 0.007 0.000 0.000 0.000 0.000 0.000 0.000 0.000
B1                0.000 0.000 0.007 0.000 0.000 0.000 0.000 0.000 0.000
C1                0.000 0.000 0.000 0.007 0.000 0.000 0.000 0.000 0.000
D1                0.000 0.000 0.000 0.000 0.007 0.000 0.000 0.000 0.000
E.L               0.000 0.000 0.000 0.000 0.000 0.021 0.000 0.000 0.000
E.Q               0.000 0.000 0.000 0.000 0.000 0.000 0.021 0.000 0.000
F.L               0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.021 0.000
F.Q               0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.021
round(det(solve(t(X)%*%X)),18)
 [1] 3e-18
```

La matrice de dispersion $(X'X)^{-1}$ du plan avec 144 essais a un déterminant plus petit que celui des deux autres plans (36 et 72). Cependant, les déterminants restent malgré tout très proche de 0 dans les 3 cas. Le compromis est donc de choisir 72 essais et de réaliser un plan $L_{72}2^4 3^2$

Exercice 5. Effets de différentes pratiques culturales sur une nouvelle variété de blé dur (exemple fictif, par S. Chauvire)

La thématique et les données sont fictives.

- ## *EXPERIENCE 1*

Une nouvelle variété de blé dur va bientôt être commercialisée. Votre rôle est de tester le rendement obtenu avec cette variété suivant les pratiques culturales utilisées. Le but est de trouver le meilleur itinéraire technique.

Liste des différentes pratiques testées :

PRATIQUE	ABREVIATION	MODALITES	
Fongicide A	A	oui	non
Fongicide B	B	oui	non

Labour	L	oui	non
Herbicide	H	oui	non
Azote	N	oui	non
Densité de semis	D	faible	forte
Date de semis	J	habituelle	précoce

1. **Combien d'essais seraient nécessaires dans le cas d'un plan complet ?**

Il y a 7 facteurs à 2 modalités donc : 2^7=128 essais.

2. **La parcelle qui vous est allouée pour ces expériences ne vous permettent de faire que 16 essais. Aux vues de ces conditions, construisez un plan adéquat.**

Pour réduire le plan à 16 essais il faut utiliser un plan 2^{7-3}. 3 facteurs seront donc confondus avec des interactions.

I	A L H B	AL AH AB LH LB HB	ALH ALB AHB LHB	ALHB
			N D J	

3. **Donnez les générateurs d'alias associés à ce plan. Quelle est la résolution du plan ?**

I=NN=*NALH* I=DD=*DALB* I=JJ=*JAHB*

Puis, en les multipliant 2 à 2 : *NDBH*, *DJLH*, *NJLB*

Et en les multipliant tous les 3 : *NDAJ*

La constante I est confondue avec des interactions d'ordre 4, la résolution du plan est donc de 4.

4. **Ce plan permet-il d'estimer les effets principaux ? les interactions d'ordre 2 ? Trouver le meilleur itinéraire technique semble-t-il réalisable ?**

Dans un plan de résolution 4, les effets principaux sont confondus avec les interactions d'ordre 3. Ces dernières pouvant facilement être considérées comme négligeables, les effets principaux seront estimés sans problème.

Les interactions d'ordre 2 sont confondues entre elles, on ne peut donc pas les estimer.

Un itinéraire technique étant une combinaison de pratiques culturales, il semble difficile d'en évaluer la valeur sans prendre en compte les interactions entre pratiques. Il faudrait plus d'essais pour les estimer.

5. **Donnez le nombre de degrés de liberté de la variance résiduelle. Qu'en déduisez-vous ?**

Il y a 16 essais donc il reste 16 ddl – 7 (facteurs) – 1 (constante) = 8 ddl.

Il serait donc possible d'estimer d'autres « paquets » d'interactions. Ainsi, une interaction d'ordre 2 pourrait être estimée SI les autres interactions avec lesquelles elle est confondue sont considérées comme négligeables.

6. **Utilisez le fonction *FrF2* du package *FrF2* pour construire le plan sous R, en ne prenant en compte que les effets linéaires. Vérifiez la qualité du plan en calculant $(X'X)^{-1}$.**

```
library(FrF2)

plan <- FrF2(nruns=16,nfactors=7,factor.names=list(A=c("oui","non"), B=c("oui","non"),
L=c("oui","non"), H=c("oui","non"), N=c("oui","non"), D=c("forte", "faible"),
J=c("precoce","habituelle")))

summary(plan)

[...]
Design generating information:
$legend
[1] A=A B=B C=L D=H E=N F=D G=J

$generators
[1] E=ABC F=ABD G=ACD


Alias structure:
$fi2
[1] AB=CE=DF AC=BE=DG AD=BF=CG AE=BC=FG AF=BD=EG AG=CD=EF BG=CF=DE


The design itself:
    A   B   L   H   N     D         J
1  non non non oui non  forte     precoce
```

```
2  non oui oui non non  forte    precoce
3  non non non non non faible habituelle
4  non non oui oui oui  forte habituelle
5  oui oui oui oui oui  forte    precoce
6  oui oui non non non faible    precoce
7  oui non non non oui  forte    precoce
8  oui non oui non non  forte habituelle
9  oui oui non oui non  forte habituelle
10 non oui non oui oui faible    precoce
11 oui oui oui non oui faible habituelle
12 non non oui non oui faible    precoce
13 oui non non oui oui faible habituelle
14 non oui oui oui non faible habituelle
15 oui non oui oui non faible    precoce
16 non oui non non oui  forte habituelle
```

```
options(contrasts=c("contr.sum","contr.sum"))
X <- model.matrix(~.,data=plan)
solve(t(X)%*%X)
```

```
            (Intercept)     A1     B1     L1     H1     N1     D1     J1
(Intercept)      0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
A1               0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
B1               0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000 0.0000
L1               0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000 0.0000
H1               0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000 0.0000
N1               0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000 0.0000
D1               0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625 0.0000
J1               0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0625
```

La matrice est diagonale, ceci confirme qu'il n'y a pas de confusion entre effets principaux.

La valeur de 0.625 = 1/16 pourrait être diminuée en augmentant le nombre d'essais.

7. **Quel modèle peut-on étudier avec ce plan ?**

Analyse de la variance à 7 facteurs.

8. **Les expériences ont été mises en place suivant un plan de ce type. Les résultats sont disponibles dans le fichier « donnees.csv ». En utilisant la fonction *AovSum* de *FactoMineR*, que déduisez-vous de ces résultats ?**

```
dta=read.csv2("donnees.csv",row.names=1)
dta$A=as.factor(dta$A)
dta$L=as.factor(dta$L)
dta$H=as.factor(dta$H)
dta$B=as.factor(dta$B)
dta$N=as.factor(dta$N)
dta$D=as.factor(dta$D)
dta$J=as.factor(dta$J)

AovSum(rdmt~A+B+L+H+N+D+J,data=dta)
```

```
Ftest
              SS df      MS F value  Pr(>F)
A         650.25  1  650.25  5.1723 0.05254 .
B         182.25  1  182.25  1.4497 0.26299
L          81.00  1   81.00  0.6443 0.44535
H         210.25  1  210.25  1.6724 0.23203
N        1225.00  1 1225.00  9.7440 0.01420 *
D           1.00  1    1.00  0.0080 0.93113
J         756.25  1  756.25  6.0154 0.03977 *
Residuals 1005.75  8  125.72
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ttest
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    64.3750     2.8031 22.9656  < 2e-16 ***
A - non        -6.3750     2.8031 -2.2743  0.05254 .
A - oui         6.3750     2.8031  2.2743  0.05254 .
B - non        -3.3750     2.8031 -1.2040  0.26299
B - oui         3.3750     2.8031  1.2040  0.26299
L - non        -2.2500     2.8031 -0.8027  0.44535
L - oui         2.2500     2.8031  0.8027  0.44535
H - non        -3.6250     2.8031 -1.2932  0.23203
H - oui         3.6250     2.8031  1.2932  0.23203
N - non         8.7500     2.8031  3.1215  0.01420 *
N - oui        -8.7500     2.8031 -3.1215  0.01420 *
D - faible     -0.2500     2.8031 -0.0892  0.93113
D - forte       0.2500     2.8031  0.0892  0.93113
J - habituelle -6.8750     2.8031 -2.4526  0.03977 *
J - precoce     6.8750     2.8031  2.4526  0.03977 *
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Le premier tableau indique les effets globaux. On voit que l’azote (N) et la date de semis (J) ont un effet significatif à 5%.

Le tableau suivant donne davantage de détails : les rendements oscillent autour de 64q/ha mais l’ajout d’azote augmente celui-ci de 8.7q/ha en moyenne et un semis précoce augmente de 6.8q/ha en moyenne.

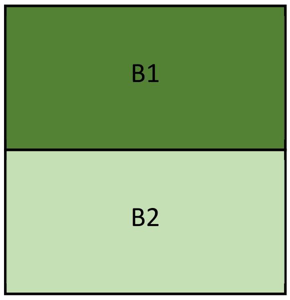
- ## *EXPERIENCE 2*

Aux vues des résultats de l’expérience précédente, vous décidez de faire une nouvelle expérience afin de préciser les effets de certains facteurs.

Les facteurs étudiés sont :

- le fongicide A (4 doses testées)
- le fongicide B (4 doses testées)
- le fongicide C (4 doses testées)
- la date de semis (2 dates testées)
- l’azote apporté (2 doses testées)

Cette fois-ci, on vous alloue une parcelle permettant de faire un grand nombre d’expériences. Cependant le terrain étant hétérogène, vous devez prendre en compte cette différence lors de la construction de votre plan d’expérience. Vous décidez de diviser la parcelle en 2 blocs plus homogènes.



9. **Un collègue vous conseille d’utiliser des carrés latins. Comment procédez-vous pour créer votre plan à l’aide de carrés latins ?**

1ère solution :

Prendre les 3 carrés latins orthogonaux d’ordre 4.

Dans le 1er carré, mettre les facteurs à 4 modalités :

	FA1	FA2	FA3	FA4
FB1	FC1	FC2	FC3	FC4
FB2	FC2	FC3	FC4	FC1
FB3	FC3	FC4	FC1	FC2
FB4	FC4	FC1	FC2	FC3

Pour le 2nd carré, transformer le facteur à 4 modalités en 2 facteurs à modalités de la manière suivante :

	N	J
1	2	2
2	2	1
3	1	2
4	1	1

On obtient alors ce 2ème carré :

	FA1		FA2		FA3		FA4	
FB1	2	2	2	1	1	2	1	1
FB2	1	2	1	1	2	2	2	1
FB3	1	1	1	2	2	1	2	2
FB4	2	1	2	2	1	1	1	2

Pour le 3ème carré latin, transformer le facteur à 4 modalités en 1 facteur à 2 modalités de la manière suivante :

	B
1	1
2	1
3	2
4	2

On obtient alors ce 3ème carré :

	FA1	FA2	FA3	FA4
FB1	1	1	2	2
FB2	2	2	1	1
FB3	1	1	2	2
FB4	2	2	1	1

Il suffit ensuite de déplier les 3 carrés obtenus.

2ème solution :

1er carré latin avec les facteurs à 4 modalités :

	FA1	FA2	FA3	FA4
FB1	FC1	FC2	FC3	FC4
FB2	FC2	FC3	FC4	FC1
FB3	FC3	FC4	FC1	FC2
FB4	FC4	FC1	FC2	FC3

Déplié :

FA	FB	FC
1	1	1
1	2	2
1	3	3
1	4	4
2	1	2
2	2	3
2	3	4
2	4	1
3	1	3
3	2	4
3	3	1
3	4	2
4	1	4
4	2	1
4	3	2
4	4	3

2ème carré latin avec les facteurs à 2 modalités :

	N1	N2
J1	B1	B2
J2	B2	B1

Déplié :

N	J	B
1	1	1
2	1	2
2	2	1
1	2	2

Il faut ensuite combiner les 2 :

	FA	FB	FC	N	J	B
exp 1	1	1	1	1	1	1
exp2	1	1	1	2	1	2
exp3	1	1	1	2	2	1
exp 4	1	1	1	1	2	2
exp 5	1	2	2	1	1	1
exp 6	1	2	2	2	1	2
exp 7	1	2	2	2	2	1
exp 8	1	2	2	1	2	2
exp 9	1	3	3	1	1	1

exp 10	1	3	3	2	1	2
exp 11	1	3	3	2	2	1
exp 12	1	3	3	1	2	2
	.	.	.	.	.	.
exp 64						

10. Un autre collègue vous suggère de passer par un plan factoriel. Comment procédez-vous ?

Pour l'instant, le plan est un plan $2^3 4^3$.

Les facteurs à 4 modalités peuvent être vus comme 2 facteurs à 2 modalités (plan 2^{9-3}).

Ex de codage permettant de retrouver le facteur à 4 modalités :

A1	A2	**FA**
1	1	**1**
1	-1	**2**
-1	1	**3**
-1	-1	**4**

Exercice 6. Trouver la meilleure recette de 'pesto' (exemple réel, par Vincent Peru)

Partie 1 : plan fractionnaire

Dans le cadre d'un projet d'ingénieur, une équipe décide de préparer une préparation culinaire de type 'pesto'. Plusieurs ingrédients ont été choisis, mais on cherche à savoir quelle combinaison offre la meilleure appréciation générale (via une note). (Thématique réelle mais simplifiée)

Les 5 ingrédients sont :

- Huile : olive ou pépin de raisin
- Fanes : radis ou carotte
- Fruit à coque : noisette ou amande
- Fromage : parmesan ou feta
- Graine : sarrasin ou lin

1. Quel est le plan complet. Quelle est la résolution d'un tel plan ?

Résolution VI, $2^5 = 32$ essais

2. On estime que seules 3 interactions pourraient être significatives : huile avec fanes, fanes avec fromage ainsi que graine avec fruit à coque. Quel est le nouveau plan d'expérience ?

Il faut au moins 9 degrés de libertés (5 facteurs + la constante + 3 interactions d'ordre 2) :

$2^{5-1} = 2^4 = 16$

3. Reste-il des interactions que nous pourrions étudier ? Si oui lesquelles ?

Degré de liberté restants : 16-9 = 7

On peut estimer toutes les interactions d'ordre 2

4. Créer ce plan sur R, et vérifier sa qualité (les 5 facteurs + les 3 interactions).

```
library(FrF2)
plan = FrF2(nruns=16, nfactors=5, factor.names=list(huile=c("olive","raisin"), fanes=c("radis","
carotte"),coque=c("noisette","amande"),fromage=c("parmesan","feta"),graine=c("sarrasin","lin")),
randomize = FALSE)
summary(plan)
```

```
Call:
FrF2(nruns = 16, nfactors = 5, factor.names = list(huile = c("olive",
    "raisin"), fanes = c("radis", "carotte"),coque = c("noisette", "amande"),
     fromage = c("parmesan", "feta"), graine = c("sarrasin", "lin")), randomize = FALSE)

Experimental design of type  FrF2
16  runs

Factor settings (scale ends):
   huile   fanes    coque  fromage   graine
```

```
1  olive   radis noisette parmesan sarrasin
2 raisin carotte   amande     feta      lin

Design generating information:
$legend
[1] A=huile   B=fanes   C=coque   D=fromage E=graine

$generators
[1] E=ABCD


Alias structure:
[[1]]
[1] no aliasing among main effects and 2fis


The design itself:
    huile   fanes    coque  fromage   graine
1   olive   radis noisette parmesan      lin
2  raisin   radis noisette parmesan sarrasin
3   olive carotte noisette parmesan sarrasin
4  raisin carotte noisette parmesan      lin
5   olive   radis   amande parmesan sarrasin
6  raisin   radis   amande parmesan      lin
7   olive carotte   amande parmesan      lin
8  raisin carotte   amande parmesan sarrasin
9   olive   radis noisette     feta sarrasin
10 raisin   radis noisette     feta      lin
11  olive carotte noisette     feta      lin
12 raisin carotte noisette     feta sarrasin
13  olive   radis   amande     feta      lin
14 raisin   radis   amande     feta sarrasin
15  olive carotte   amande     feta sarrasin
16 raisin carotte   amande     feta      lin
class=design, type= FrF2

options(contrasts=c("contr.sum","contr.sum"))
X = model.matrix(~huile+fanes+coque+fromage+graine+huile:fanes+fanes:fromage+coque:graine, data=
plan)
solve(t(X)%*%X)

                (Intercept) huile1 fanes1 coque1 fromage1 graine1 huile1:fanes1
(Intercept)          0.0625 0.0000 0.0000 0.0000   0.0000  0.0000        0.0000
huile1               0.0000 0.0625 0.0000 0.0000   0.0000  0.0000        0.0000
fanes1               0.0000 0.0000 0.0625 0.0000   0.0000  0.0000        0.0000
coque1               0.0000 0.0000 0.0000 0.0625   0.0000  0.0000        0.0000
fromage1             0.0000 0.0000 0.0000 0.0000   0.0625  0.0000        0.0000
graine1              0.0000 0.0000 0.0000 0.0000   0.0000  0.0625        0.0000
huile1:fanes1        0.0000 0.0000 0.0000 0.0000   0.0000  0.0000        0.0625
fanes1:fromage1      0.0000 0.0000 0.0000 0.0000   0.0000  0.0000        0.0000
coque1:graine1       0.0000 0.0000 0.0000 0.0000   0.0000  0.0000        0.0000
                fanes1:fromage1 coque1:graine1
(Intercept)              0.0000         0.0000
huile1                   0.0000         0.0000
fanes1                   0.0000         0.0000
coque1                   0.0000         0.0000
fromage1                 0.0000         0.0000
graine1                  0.0000         0.0000
huile1:fanes1            0.0000         0.0000
fanes1:fromage1          0.0625         0.0000
coque1:graine1           0.0000         0.0625
```

Il y a orthogonalité de tous les effets entre eux, et de tous les effets avec les interactions, et des interactions entre elles. C'est donc un plan de qualité.

<u>Partie 2 : plan optimal</u>

On s'interroge sur la possibilité de faire une recette végétale. Pour cela, on rajoute à nos essais une possibilité sans fromage. Pour des raisons de budget limité, on ne peut faire que 12 essais et ne faire tester qu'à 5 juges (les membres de l'équipe).

5. **Donner un plan d'expérience sur R et commenter sa qualité.**

4 facteurs à 2 (huile, fanes, fruit à coque, graine), 1 à 3 (fromage : parmesan/feta/rien) et 1 à 5 (juge)

```
library(DoE.base)
library(AlgDesign)

data = fac.design(nlevels=c(2,2,2,2,3,5), seed=1234)
des = optFederov(~.,data=data,12)
```

```
X = model.matrix(~., data=des$design)
round(solve(t(X)%*%X),digits = 2)

            (Intercept)    A1    B1    C1    D1   E.L   E.Q   F.L   F.Q   F.C   F^4
(Intercept)        0.09  0.01  0.00  0.01  0.00  0.00  0.01  0.01  0.00  0.02  0.01
A1                 0.01  0.10 -0.02  0.00 -0.02 -0.04  0.02 -0.02  0.00 -0.02 -0.03
B1                 0.00 -0.02  0.09  0.02  0.00  0.01  0.00  0.03 -0.02 -0.01 -0.02
C1                 0.01  0.00  0.02  0.10  0.02  0.04  0.02  0.01  0.03 -0.03 -0.01
D1                 0.00 -0.02  0.00  0.02  0.09  0.01  0.00 -0.02  0.03  0.01  0.02
E.L                0.00 -0.04  0.01  0.04  0.01  0.32  0.00  0.11  0.09 -0.04  0.07
E.Q                0.01  0.02  0.00  0.02  0.00  0.00  0.30  0.01  0.08  0.04 -0.11
F.L                0.01 -0.02  0.03  0.01 -0.02  0.11  0.01  0.47  0.09 -0.01  0.05
F.Q                0.00  0.00 -0.02  0.03  0.03  0.09  0.08  0.09  0.51  0.00 -0.04
F.C                0.02 -0.02 -0.01 -0.03  0.01 -0.04  0.04 -0.01  0.00  0.45 -0.06
F^4                0.01 -0.03 -0.02 -0.01  0.02  0.07 -0.11  0.05 -0.04 -0.06  0.53
```

Il y a une légère confusion entre fromage et juge, mais sinon on observe que les valeurs les plus élevées sont sur la diagonale, donc on peut considérer que c'est un plan acceptable.

Exercice 7. Élaboration de quatre quarts (exemple réel, par Emma Rouault)

Partie 1. Tri des effets des facteurs

Des étudiants de DUT génie biologique option industries agroalimentaires et biologiques (IAB) participent à un concours culinaire. Ils souhaitent remporter ce concours grâce à l'élaboration d'un quatre quarts. Leur but est de s'intéresser à l'effet de chaque ingrédient du quatre quarts sur la recette finale.

Un quatre quarts est fabriqué de farine, de sucre (on utilisera pour la recette du saccharose industriel), des œufs, de la levure et du beurre.

On considère qu'il existe deux mesures possibles pour chaque ingrédient :

- Farine : 40 – 50
- Saccharose : 40 – 50
- Blanc œuf : 0 – 20
- Jaune œuf : 0 – 20
- Beurre : 50 – 54
- Levure : 1 – 3

1. **Que proposeriez-vous comme plan afin d'aider ces étudiants ?**

On s'intéresse à l'effet de 6 facteurs (x1,x2,x3,x4,x5 et x6) sur une recette finale de quatre quart.

On veut donc savoir si les variables X explicatives (x1,x2,x3,x4,x5 et x6) ont un effet sur la variable réponse Y qui est la recette finale. Dans notre cas, on étudie l'effet de variables X quantitatives sur une variable Y quantitative.

On propose tout d'abord d'élaborer un plan factoriel complet afin de connaître le nombre d'essais à réaliser pour avoir une orthogonalité parfaite des différents facteurs.

2. **Quel est le nombre d'essais associés au plan complet ?**

Ici, on a 6 facteurs à 2 modalités chacun.

Le plan complet est donc $2^6 = 64$ essais. En 64 essais, on est sûr d'étudier tous les facteurs avec toutes leurs modalités.

3. **Finalement, les étudiants rencontrent des difficultés pour réaliser leurs quatre quarts et s'interrogent maintenant sur le nombre minimum d'essais à réaliser pour faire leurs quatre quarts. Quel nombre d'essais doivent impérativement être réalisés au minimum ?**

$1+6 \times (2-1) = 7$ essais. Car il y a 6 facteurs à 2 modalités et 1 degré de liberté pour la constante. 7 essais minimum sont possibles mais dans ce cas, on perd l'orthogonalité.

4. **Les étudiants peuvent mettre seulement 7 quatre quarts dans un four industriel en même temps. Selon votre réponse à la question précédente, en combien de fournées les quatre quarts seront prêts ?**

Les étudiants pourront réaliser le nombre d'essais minimum en seulement une fournée (car il y a 7 essais minimum et on peut mettre 7 quatre quarts dans un four industriel).

En théorique, cela marcherait. Mais en pratique, il faut aussi faire attention de tenir compte de l'emplacement du quatre quarts dans le four. Cela peut jouer dans la variabilité du rendement final de la recette.

5. **Construire le plan complet sur R puis le plan optimal en 14 essais.**

<u>Construction du plan complet</u>

```
set.seed(123)
library(DoE.base)
plan <- fac.design(nlevels=2, nfactors=6,factor.names=c("Farine","Saccharose","Blanc
oeuf","Jaune oeuf","Beurre","Levure"),randomize=FALSE)
plan
```

On utilise set.seed pour générer les mêmes valeurs à chaque fois afin de pouvoir obtenir un code reproductible lorsque l'on fait appel à un processus aléatoire.

On utilise randomize = FALSE pour donner les essais toujours dans le même ordre.

Avec la fonction fac.design on a le plan complet, en 64 essais. Dans ce plan, il est proposé aux étudiants de réaliser 64 recettes de quatre quarts afin d'étudier tous les effets de chaque ingrédient. Pour chaque ingrédient, il est indiqué s'il prend sa modalité 1 ou 2, pour la réalisation de la recette. Par exemple, si on regarde pour la recette n°2, R nous donne :

```
   Farine Saccharose Blanc.oeuf Jaune.oeuf Beurre Levure
1       1          1          1          1      1      1
2       2          1          1          1      1      1
3       1          2          1          1      1      1
4       2          2          1          1      1      1
5       1          1          2          1      1      1
```

Cela veut donc dire que l'essai n°2 devra être composée des modalités 1 de tous les ingrédients sauf la farine qui prendra la modalité 2. Ainsi, l'essai n°2 sera composé de :

- 50g de farine
- 40 g de Saccharose
- 0 g de blanc d'œuf
- 0 g de jaune d'œuf
- 50 g de beurre
- 1 g de levure

Ce modèle est le meilleur, il permet d'avoir une parfaite orthogonalité entre les facteurs. Afin d'élire la meilleure recette, il faudrait faire une analyse sensorielle avec un jury expérimenté pour qu'ils choisissent le quatre quarts à présenter à leurs concours.

<u>Construction du plan optimal en 14 essais</u>

```
library(AlgDesign)
planOpt <- optFederov(~.,plan,nTrials=14,criterion="D",nRepeats = 20)
planOpt
```

Le nouveau modèle prend en compte tous les facteurs de la base « plan » et on dit qu'on veut faire 14 essais au total (nTrials=14) et répéter l'algorithme 20 fois (nRepeats=20) pour trouver le meilleur plan.

La fonction optFederov va nous donner un plan qui minimise le déterminant de la matrice de covariance. On aura alors un plan qui permet d'estimer au mieux les paramètres du modèle.

L'algorithme nous a donc choisi 14 recettes à réaliser qui sont aussi disponibles dans le plan complet (il s'agit des essais n°6,7,10,12,19,25,32,33,43,44,45,50,56 et 61) : (la ligne $rows de la sortie R nous récapitule les numéros de ces essais.

```
$design
   Farine Saccharose Blanc.oeuf Jaune.oeuf Beurre Levure
6       2          1          2          1      1      1
7       1          2          2          1      1      1
10      2          1          1          2      1      1
12      2          2          1          2      1      1
19      1          2          1          1      2      1
25      1          1          1          2      2      1
32      2          2          2          2      2      1
33      1          1          1          1      1      2
43      1          2          1          2      1      2
44      2          2          1          2      1      2
45      1          1          2          2      1      2
50      2          1          1          1      2      2
56      2          2          2          1      2      2
61      1          1          2          2      2      2
```

```
$rows
 [1]  6  7 10 12 19 25 32 33 43 44 45 50 56 61
```

Par exemple, pour réaliser l'essai n°6 on fera la recette suivante :

- 50g de farine (modalité 2)
- 40 g de Saccharose (modalité 1)
- 20 g de blanc d'œuf (modalité 2)
- 0 g de jaune d'œuf (modalité 1)
- 50 g de beurre (modalité 1)
- 1 g de levure (modalité 1)

6. Vérifier la qualité des plans

Vérification de la qualité du plan complet

```
X<-model.matrix(~.,plan)
t(X)%*%X
solve(t(X)%*%X)
```

```
            (Intercept) Farine1 Saccharose1 Blanc.oeuf1 Jaune.oeuf1 Beurre1 Levure1
(Intercept)          64       0           0           0           0       0       0
Farine1               0      64           0           0           0       0       0
Saccharose1           0       0          64           0           0       0       0
Blanc.oeuf1           0       0           0          64           0       0       0
Jaune.oeuf1           0       0           0           0          64       0       0
Beurre1               0       0           0           0           0      64       0
Levure1               0       0           0           0           0       0      64
```

```
            (Intercept)  Farine1 Saccharose1 Blanc.oeuf1 Jaune.oeuf1  Beurre1  Levure1
(Intercept)    0.015625 0.000000    0.000000    0.000000    0.000000 0.000000 0.000000
Farine1        0.000000 0.015625    0.000000    0.000000    0.000000 0.000000 0.000000
Saccharose1    0.000000 0.000000    0.015625    0.000000    0.000000 0.000000 0.000000
Blanc.oeuf1    0.000000 0.000000    0.000000    0.015625    0.000000 0.000000 0.000000
Jaune.oeuf1    0.000000 0.000000    0.000000    0.000000    0.015625 0.000000 0.000000
Beurre1        0.000000 0.000000    0.000000    0.000000    0.000000 0.015625 0.000000
Levure1        0.000000 0.000000    0.000000    0.000000    0.000000 0.000000 0.015625
```

Ce plan est parfait. On peut voir grâce à la matrice inverse $(X'X)^{-1}$ (solve(t(X)%*%X)) que les effets principaux sont orthogonaux et on trouve 1/64 = 0.015625 sur la diagonale de la matrice.

Vérification de la qualité du plan optimal en 14 essais

```
X1<-model.matrix(~.,planOpt$design)
t(X1)%*%X1
round(solve(t(X1)%*%X1),4)
```

On utilise $design pour avoir le plan d'expérience.

```
            (Intercept) Farine1 Saccharose1 Blanc.oeuf1 Jaune.oeuf1 Beurre1 Levure1
(Intercept)          14       0           0          -2           2      -2       0
Farine1               0      14           2           0           0       0      -2
Saccharose1           0       2          14           0           0       0      -2
Blanc.oeuf1          -2       0           0          14          -2       2       0
Jaune.oeuf1           2       0           0          -2          14      -2       0
Beurre1              -2       0           0           2          -2      14       0
Levure1               0      -2          -2           0           0       0      14
```

```
            (Intercept) Farine1 Saccharose1 Blanc.oeuf1 Jaune.oeuf1 Beurre1 Levure1
(Intercept)      0.0750  0.0000      0.0000      0.0083     -0.0083  0.0083  0.0000
Farine1          0.0000  0.0741     -0.0093      0.0000      0.0000  0.0000  0.0093
Saccharose1      0.0000 -0.0093      0.0741      0.0000      0.0000  0.0000  0.0093
Blanc.oeuf1      0.0083  0.0000      0.0000      0.0750      0.0083 -0.0083  0.0000
Jaune.oeuf1     -0.0083  0.0000      0.0000      0.0083      0.0750  0.0083  0.0000
Beurre1          0.0083  0.0000      0.0000     -0.0083      0.0083  0.0750  0.0000
Levure1          0.0000  0.0093      0.0093      0.0000      0.0000  0.0000  0.0741
```

Ce plan n'est pas parfait, il n'y a pas orthogonalité parfaite entre les facteurs.

On observe une confusion entre les facteurs, la matrice inversée n'est pas diagonale.

Par exemple, l'effet du beurre n'est pas dû que à la quantité de beurre mais aussi à la quantité de blanc d'œuf et de jaune d'œuf.

Partie 2. Étude des effets des facteurs sur le gonflement d'un quatre quart

Les mêmes étudiants de l'IUT de l'exercice précédent veulent maintenant s'intéresser à l'effet de la quantité de jaune d'œuf, de blanc d'œuf, de levure ainsi qu'à l'effet de l'opérateur sur le gonflement des quatre quarts. On considère qu'il y a seulement deux étudiants du groupe participant au concours culinaire qui réalisent les différentes recettes.

On considère toujours qu'il existe deux modalités possibles :

- Jaune œuf : faible (quantité) – raisonnable (quantité)
- Blanc œuf : faible (quantité) – raisonnable (quantité)
- Levure : sans (levure) – avec (levure)
- Opérateurs : A - B

On considère également que la modalité 1 correspond pour le jaune d'œuf et le blanc d'œuf à « faible», pour la levure à « sans » et pour l'opérateur à « A ».

Concernant la modalité 2, elle correspond à « raisonnable » pour le jaune d'œuf et le blanc d'œuf, à « avec » pour la levure et à « B » pour l'opérateur.

Les résultats de l'expérience sont les suivantes (matrice des essais) :

Essai	Jaune d'œuf	Blanc d'œuf	Levure	Opérateur	Gonflement
1	1	2	1	2	29.03
2	2	2	1	1	39.18
3	1	2	2	1	33.44
4	1	1	1	1	24.24
5	2	2	2	2	34.36
6	2	1	2	1	37.03
7	2	1	1	2	37.13
8	1	1	2	2	38.04

7. **Construire la matrice des effets correspondante en prenant compte que la modalité 1 prend pour valeur « 1 » et que la modalité 2 prend pour valeur « -1 »**

Jaune d'œuf	Blanc d'œuf	Levure	Opérateur/Interaction Jauneoeuf,Blancoeuf,Levure
1	-1	1	-1
-1	-1	1	1
1	-1	-1	1
1	1	1	1
-1	-1	-1	-1
-1	1	-1	1
-1	1	1	-1
1	1	-1	-1

8. **Comment appelle-t-on un tel type de plan ? Les effets principaux sont-ils orthogonaux ?**

On étudie l'effet de variables qualitatives sur une réponse Y quantitative (Gonflement).

Il s'agit d'un plan fractionnaire classique 2^{4-1}. Les 3 premiers facteurs (Jaune d'œuf, Blanc d'œuf, Levure) constituent un plan complet (2^3=8 essais) et le quatrième facteur (opérateur) est confondu avec l'interaction des 3 premiers facteurs. Le plan étant un plan fractionnaire, on s'assure ici que les effets principaux sont tous orthogonaux.

9. **Étudiez ces données sur R et en conclure sur les effets de chaque facteur sur le gonflement du quatre quarts.**

```
library(FactoMineR)
Jauneoeuf <- as.factor(c(1,2,1,1,2,2,2,1))
```

```
Blancoeuf <- as.factor(c(2,2,2,1,2,1,1,1))
Levure <- as.factor(c(1,1,2,1,2,2,1,2))
Operateur <- as.factor(c(2,1,1,1,2,1,2,2))
Gonflement <- c(29.03,24.24,30,39.18,34.36,40,37.13,38.04)
test <- cbind.data.frame(Jauneoeuf,Blancoeuf,Levure,Operateur)
AovSum(Gonflement~Jauneoeuf+Blancoeuf+Levure+Operateur,data=test)
```

Tout d'abord, on attribue à chaque facteur sa modalité (récupérée dans la matrice des essais donnée en énoncé) grâce à la fonction as.factor. Ensuite, on attribue les valeurs prises par la réponse Y (ici, le gonflement) lors des essais des étudiants.

On utilise ensuite la fonction cbind pour mettre toutes les variables explicatives X ainsi que la réponse Y dans un tableau. On « colle » ainsi les facteurs les uns à côté des autres.

Puis, à l'aide de la fonction AoVSum, on obtient tout d'abord un tableau de fisher pour savoir s'il y a un effet d'un ou de plusieurs facteurs sur le gonflement. Puis AovSum nous donne ensuite un tableau de student pour savoir parmi tous les facteurs, quels sont ceux (en précisant leurs modalité) qui ont un effet sur le gonflement (Y).

```
Ftest
               SS df      MS F value   Pr(>F)  
Jauneoeuf   0.034  1   0.034  0.0031 0.95904  
Blancoeuf 168.545  1 168.545 15.5037 0.02919 *
Levure     20.544  1  20.544  1.8898 0.26291  
Operateur   3.302  1   3.302  0.3038 0.61992  
Residuals  32.614  3  10.871                  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Ttest
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)    33.9975     1.1657 29.1643    9e-05 ***
Jauneoeuf - 1   0.0650     1.1657  0.0558  0.95904    
Jauneoeuf - 2  -0.0650     1.1657 -0.0558  0.95904    
Blancoeuf - 1   4.5900     1.1657  3.9375  0.02919 *  
Blancoeuf - 2  -4.5900     1.1657 -3.9375  0.02919 *  
Levure - 1     -1.6025     1.1657 -1.3747  0.26291    
Levure - 2      1.6025     1.1657  1.3747  0.26291    
Operateur - 1  -0.6425     1.1657 -0.5512  0.61992    
Operateur - 2   0.6425     1.1657  0.5512  0.61992    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

En étudiant les résultats, on se rend compte que le seul facteur ayant un effet sur le gonflement est le blanc d'œuf, qu'il prenne la modalité 1 ou 2 (quantité faible ou raisonnable).

En effet, c'est le seul effet significatif (pvalue = 0.02919 < 0.05).

10. Étudiez la qualité de ce modèle

```
options(contrasts=c("contr.sum","contr.sum")) #permet de dire que l'on veut construire une matrice x avec une condition : somme des alpha i=0
X <- model.matrix(~.,data=test)
t(X)%*%X
solve(t(X)%*%X)
```

On utilise la première ligne de code pour expliquer que l'on veut construire une matrice X avec une condition : $\sum \alpha_i=0$

```
            (Intercept) Jauneoeuf1 Blancoeuf1 Levure1 Operateur1
(Intercept)           8          0          0       0          0
Jauneoeuf1            0          8          0       0          0
Blancoeuf1            0          0          8       0          0
Levure1               0          0          0       8          0
Operateur1            0          0          0       0          8

            (Intercept) Jauneoeuf1 Blancoeuf1 Levure1 Operateur1
(Intercept)       0.125      0.000      0.000   0.000      0.000
Jauneoeuf1        0.000      0.125      0.000   0.000      0.000
Blancoeuf1        0.000      0.000      0.125   0.000      0.000
Levure1           0.000      0.000      0.000   0.125      0.000
Operateur1        0.000      0.000      0.000   0.000      0.125
```

Grâce à la matrice inverse $(X'X)^{-1}$ (solve(t(X)%*%X)), on voit que les effets principaux sont orthogonaux et on trouve 1/8 = 0.0125 sur la diagonale de la matrice. On ne peut pas faire mieux comme plan. Les paramètres sont tous très bien estimés et comme dit précédemment la variance est faible (0.125), et les covariances sont nulle ce qui marque l'indépendance des essais.

Exercice 8. Expériences autour de l'acide lactobionique (problématique réelle, par Chloé Derouet)

L'acide lactobionique est un coproduit du lactose qui est de plus en plus étudié. Son utilisation peut être multiple que ce soit en majorité au sein des produits pharmaceutiques mais aussi au sein de l'alimentation. Ainsi, un laboratoire de recherche a décidé d'étudier cet acide lactobionique et les réactions de formation de ce composé. En effet, le laboratoire essaie de trouver la meilleure combinaison de facteur afin de former la plus grande quantité d'acide lactobionique.

Lors d'essais préliminaire, il a été constaté que 5 facteurs étaient susceptibles de s'influencer entre eux et d'influencer la quantité d'acide lactobionique formée. Ainsi, il a été décidé d'étudier 5 facteurs à deux niveaux présentés dans la liste suivante :

- A : La quantité de lactose : 0,1g ;0,25g
- B : Le pH de la solution : 4,7 ; 6,7
- C : L'enzyme utilisée : Lactoyield, laccase
- D : Tube fermé : oui,non
- E : La T° : 25°C, 30°C

Les informations sont issues d'un stage réalisé au cours du S7 au Canada.

Partie n°1 : Construction d'un plan afin de trouver la meilleure combinaison pour la production d'acide lactobionique

1. **Pour commencer, si le laboratoire ne se donne pas de nombre d'essais maximal à réaliser, combien d'essais doivent être réalisés en cas d'utilisation du plan complet ?**

Le laboratoire va ainsi étudier un plan complet à 5 facteurs à 2 niveaux ce qui correspond à 2^5= 32 essais.

2. **Le problème est que le les enzymes coûtent assez cher ce qui pousse le laboratoire à ne réaliser au maximum que 16 essais, ainsi quel sera le plan de base à construire, construisez la matrice des effets de ce plan de base pour le modèle complet (i.e. avec toutes les interactions possibles) ?**

Si le nombre d'essais est de 16 et que le nombre de facteur est de 5 avec 2 modalités chacun alors le plan sera : 2^{5-1}.

La matrice des effets est :

I	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1
1	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1
1	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1
1	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1
1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1
1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1
1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1
1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1
1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1
1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1
1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1
1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1
1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1
1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1

3. **Maintenant que vous avez construit la matrice des effets, avec quels effets souhaitez-vous confondre le facteur T° ? Quel est la résolution du plan ?**

Ce plan possède 4 facteurs mais de base nous voulions étudier 5 facteurs alors il faut confondre le facteur E avec l'interaction d'ordre le plus élevé c'est-à-dire ABCD. Ainsi, le générateur d'alias est I est confondu avec ABCDE. La résolution du plan est donc

V car le plus petit générateur d'alias est ABCDE. En résolution V, les facteurs principaux sont confondus avec des interactions d'ordre 4 ou plus et les interactions (d'ordre 2) sont confondues avec des interactions d'ordre 3 ou plus.

4. **Construisez le plan sur R et analysez l'efficacité du plan**

Construction du plan :

```
plan1<-FrF2(nruns=16,nfactors = 5,factor.names = list(qt_la=c("0,1","0,25"),pH=c("4,7","6,7"),enzyme=c("LY","laccase"),tube_fermé=c("oui","non"),T=c("25","30")))
summary(plan1):
Call:
FrF2(nruns = 16, nfactors = 5, factor.names = list(qt_la = c("0,1",
    "0,25"), pH = c("4,7", "6,7"), enzyme = c("LY",
    "laccase"), tube_fermé = c("oui", "non"),
    T = c("25", "30")))

Experimental design of type  FrF2
16  runs

Factor settings (scale ends):
  qt_la  pH  enzyme tube_fermé  T
1   0,1 4,7      LY        oui 25
2  0,25 6,7 laccase        non 30

Design generating information:
$legend
[1] A=qt_la      B=pH         C=enzyme     D=tube_fermé E=T

$generators
[1] E=ABCD

Alias structure:
[[1]]
[1] no aliasing among main effects and 2fis

The design itself:
   qt_la  pH  enzyme tube_fermé  T
1    0,1 4,7      LY        non 25
2    0,1 6,7 laccase        oui 30
3   0,25 6,7      LY        non 25
4    0,1 4,7 laccase        oui 25
5   0,25 4,7      LY        non 30
6   0,25 6,7      LY        oui 30
7    0,1 6,7 laccase        non 25
8   0,25 4,7      LY        oui 25
9   0,25 6,7 laccase        oui 25
10   0,1 4,7 laccase        non 30
11  0,25 6,7 laccase        non 30
12  0,25 4,7 laccase        oui 30
13   0,1 6,7      LY        non 30
14  0,25 4,7 laccase        non 25
15   0,1 6,7      LY        oui 25
16   0,1 4,7      LY        oui 30
class=design, type= FrF2
```

La sortie R permet de voir que le générateur d'alias est E=ABCD ce qui correspond à ce que j'ai dis à la question précédente avec E confondu avec ABCD et I confondu avec ABCDE, nous avons bien un plan de résolution 5. Ce plan nous permet donc d'étudier les facteurs principaux mais aussi toutes les interactions sans biais.

Efficacité du plan :

```
X2<-model.matrix(~qt_la+pH+enzyme+tube_fermé+T+qt_la:pH+qt_la:enzyme+qt_la:tube_fermé+qt_la:T+pH:enzyme+pH:tube_fermé+pH:T+enzyme:tube_fermé+enzyme:T+tube_fermé:T,data = plan1)
t(X2)%*%X2
                   (Intercept) qt_la1 pH1 enzyme1 tube_fermé1 T1 qt_la1:pH1 qt_la1:enzyme1 qt_la1:tube_fermé1 qt_la1:T1 pH1:enzyme1
(Intercept)                 16      0   0       0           0  0          0              0                  0         0           0
qt_la1                       0     16   0       0           0  0          0              0                  0         0           0
pH1                          0      0  16       0           0  0          0              0                  0         0           0
enzyme1                      0      0   0      16           0  0          0              0                  0         0           0
tube_fermé1                  0      0   0       0          16  0          0              0                  0         0           0
T1                           0      0   0       0           0 16          0              0                  0         0           0
qt_la1:pH1                   0      0   0       0           0  0         16              0                  0         0           0
qt_la1:enzyme1               0      0   0       0           0  0          0             16                  0         0           0
qt_la1:tube_fermé1           0      0   0       0           0  0          0              0                 16         0           0
qt_la1:T1                    0      0   0       0           0  0          0              0                  0        16           0
pH1:enzyme1                  0      0   0       0           0  0          0              0                  0         0          16
pH1:tube_fermé1              0      0   0       0           0  0          0              0                  0         0           0
pH1:T1                       0      0   0       0           0  0          0              0                  0         0           0
enzyme1:tube_fermé1          0      0   0       0           0  0          0              0                  0         0           0
enzyme1:T1                   0      0   0       0           0  0          0              0                  0         0           0
tube_fermé1:T1               0      0   0       0           0  0          0              0                  0         0           0
            pH1:tube_fermé1 pH1:T1 enzyme1:tube_fermé1 enzyme1:T1 tube_fermé1:T1
(Intercept)               0      0                   0          0              0
qt_la1                    0      0                   0          0              0
pH1                       0      0                   0          0              0
enzyme1                   0      0                   0          0              0
```

```
tube_fermé1              0    0           0    0     0
T1                       0    0           0    0     0
qt_la1:pH1               0    0           0    0     0
qt_la1:enzyme1           0    0           0    0     0
qt_la1:tube_fermé1       0    0           0    0     0
qt_la1:T1                0    0           0    0     0
pH1:enzyme1              0    0           0    0     0
pH1:tube_fermé1         16    0           0    0     0
pH1:T1                   0   16           0    0     0
enzyme1:tube_fermé1      0    0          16    0     0
enzyme1:T1               0    0           0   16     0
tube_fermé1:T1           0    0           0    0    16
```

On constate que les essais sont bien choisis car la matrice est orthogonale.

5. **Que se passe-t-il si le laboratoire veut diminuer le nombre d'essai à 8 afin d'économiser? En cas de passage à 8 essais comment analyserait-il le plan ? Le laboratoire doit-il diminuer le nombre d'essai ?**

Le laboratoire s'était donné maximum 16 essais à réaliser, ces 16 essais permettent avec un plan de résolution V d'étudier toutes les interactions sans biais. Si le laboratoire souhaite passer à 8 essais nous passerions donc à un plan 2^{5-2} . Il faudrait donc confondre 2 facteurs avec 2 effets du type AB,AC. Ainsi la plus petite interaction avec laquelle l'interaction sera confondue sera d'ordre 3 du type ABD ou ABC.

Nous aurions donc un plan de résolution III ce qui fait que seul les effets principaux seront estimés et les interactions d'ordre 2 ou plus seront considérées comme négligeables.

L'analyse des résultats du plan avec 8 essais se ferait avec une analyse de variance à 5 facteurs sans interactions.

La formule R serait : `AovSum(Y~qt_la+pH+enzyme+tube_fermé+T+,data = plan1)`

Avec Y qui représente les résultats c'est-à-dire la concentration d'acide lactobionique.

Ce plan a 8 essais n'est donc pas souhaitable par le laboratoire car il est nécessaire qu'il puisse étudier les interactions entre les facteurs car au cours d'une réaction les interactions sont importantes.

Partie n°2 : Construction d'un plan optimal

Comme pour la première part, on s'intéresse à l'effet de 5 facteurs à 2 modalités sur une réponse Y :

- A : La quantité de lactose : 0,1g ;0,25g
- B : Le pH de la solution : 4,7 ; 6,7
- C : L'enzyme utilisée : Lactoyield, laccase
- D : Tube fermé : oui,non
- E : La T° : 25°C, 30°C

6. **Pour commencer reconstruisez le plan complet mais avec cette fois ci la fonction fac.design puis construisez le plan complet à 16 essais avec la fonction optFederov du package AlgDesign**

<u>Construction du plan complet :</u>

```
library(DoE.base)
plan2<-fac.design(nlevels = 2,nfactors = 5)
plan2
   A B C D E
1  1 2 1 2 1
2  2 2 1 2 1
3  1 2 2 2 2
4  1 1 2 1 1
5  1 2 2 1 2
6  1 1 2 2 2
7  2 2 1 1 1
8  1 1 2 1 2
9  2 1 1 1 2
10 2 2 2 1 2
11 2 1 2 1 2
12 2 1 2 2 1
13 1 2 1 1 1
14 2 1 1 2 2
15 2 1 2 2 2
16 2 1 2 1 1
17 1 2 2 2 1
18 1 2 1 2 2
19 2 2 1 2 2
20 1 1 2 2 1
21 2 2 2 1 1
22 2 2 2 2 1
23 2 2 1 1 2
24 1 1 1 1 1
25 1 1 1 2 1
26 1 2 1 1 2
27 2 1 1 1 1
28 2 2 2 2 2
29 2 1 1 2 1
30 1 2 2 1 1
31 1 1 1 1 2
32 1 1 1 2 2
```

Nous avons bien 32 essais dans ce plan complet.

<u>Construction du plan optimal en 16 essais :</u>

```
library(AlgDesign)
planOpt<-optFederov(~.,planX,nTrials = 16,criterion = "D")
planOpt
$D
[1] 1

$A
[1] 1

$Ge
[1] 1

$Dea
[1] 1

$design
   A B C D E
3  1 2 2 2 2
5  1 2 2 1 2
7  2 2 1 1 1
8  1 1 2 1 2
11 2 1 2 1 2
12 2 1 2 2 1
13 1 2 1 1 1
14 2 1 1 2 2
16 2 1 2 1 1
17 1 2 2 2 1
19 2 2 1 2 2
22 2 2 2 2 1
23 2 2 1 1 2
24 1 1 1 1 1
25 1 1 1 2 1
32 1 1 1 2 2

$rows
 [1]  3  5  7  8 11 12 13 14 16 17 19 22 23 24 25 32
```

7. **Calculer la matrice des effets associée à votre modèle et votre plan et calculer ensuite la matrice d'information ? Que constatez-vous ?**

```
options(contrasts = c("contr.sum","contr.sum"))
Xopti<-model.matrix(~.,planOpt$design)
t(Xopti)%*%Xopti
            (Intercept) A1 B1 C1 D1 E1
(Intercept)          16  0  0  0  0  0
A1                    0 16  0  0  0  0
B1                    0  0 16  0  0  0
C1                    0  0  0 16  0  0
D1                    0  0  0  0 16  0
E1                    0  0  0  0  0 16
```

On constate que le plan est orthogonal car nous n'avons que des 16 sur la diagonale et des 0 partout ailleurs. On remarque aussi que nous avons le même résultat que si nous avions calculé la matrice d'information du plan fractionnaire 2^{5-1}.

Maintenant le laboratoire constate que certains facteurs peuvent être définis selon d'autres modalités. Voici les changement transmis par le labo :

- A : La quantité de lactose : 0,1g ;0,25g ;0,3g
- B : Le pH de la solution : 4,7 ; 6,7
- C : L'enzyme utilisée : Lactoyield, laccase
- D : Tube fermé : oui,non
- E : La T° :20°C, 25°C, 30°C

Ainsi, les facteurs B,C et D ont 2 modalités tandis que les facteurs A et E ont 3 modalités

8. **Quels est le nombre d'essai associé au plan complet et construisez le ? Combien d'essais avons-nous besoins si on souhaite l'orthogonalité entre les facteurs 2 à 2 ?**

Nombre d'essai du plan complet : $2^3*3^2=72$ essais

PPCM(6,8,9)=72, ainsi le plan orthogonal nécessite autant d'essais que pour le plan complet.

9. **Construisez un plan optimal en 16 essais et calculez la matrice des effets associée à votre modèle et votre plan, puis calculez la matrice de dispersion, commentez ce plan.**

```
Design2<-fac.design(nlevels = c(3,2,2,2,3), factor.names = c("A","B","C","D","E"))
plan.2opt<-optFederov(~.,Design2,nTrials = 16,criterion = "D")
Xopti2=model.matrix(~.,plan.2opt$design)
VARCOV<-solve(t(Xopti2)%*%Xopti2)
round(VARCOV,2)

             (Intercept)   A.L   A.Q    B1    C1    D1   E.L   E.Q
(Intercept)         0.07  0.02  0.01  0.00  0.00  0.00  0.01 -0.01
A.L                 0.02  0.21  0.03  0.00  0.00  0.00  0.02 -0.01
A.Q                 0.01  0.03  0.18  0.00  0.00  0.00  0.01 -0.01
B1                  0.00  0.00  0.00  0.06  0.00  0.00 -0.01 -0.02
C1                  0.00  0.00  0.00  0.00  0.06  0.00 -0.01 -0.02
D1                  0.00  0.00  0.00  0.00  0.00  0.06 -0.01 -0.02
E.L                 0.01  0.02  0.01 -0.01 -0.01 -0.01  0.19  0.01
E.Q                -0.01 -0.01 -0.01 -0.02 -0.02 -0.02  0.01  0.21
plan.2opt$D
[1] 0.5656461
```

On peut constater que la matrice n'est pas complétement orthogonale. Mais le coefficient D nous permet de voir que nous avons une efficacité de 56,5% pour l'estimation de chaque facteur par rapport au plan complet.

10. **Pour finir, construisez la matrice de dispersion du plan en 72 essais pour vérifier son orthogonalité**

```
Design.1.Dopt <- optFederov(~.,Design2,nTrials=72,criterion="D")
Xopti <- model.matrix(~., Design.1.Dopt$design)
VARCOVopti <- solve(t(Xopti)%*%Xopti)
round(VARCOVopti,
            (Intercept)  A.L  A.Q   B1   C1   D1  E.L  E.Q
(Intercept)        0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00
A.L                0.00 0.04 0.00 0.00 0.00 0.00 0.00 0.00
A.Q                0.00 0.00 0.04 0.00 0.00 0.00 0.00 0.00
B1                 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00
C1                 0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.00
D1                 0.00 0.00 0.00 0.00 0.00 0.01 0.00 0.00
E.L                0.00 0.00 0.00 0.00 0.00 0.00 0.04 0.00
E.Q                0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.04
```

On trouve bien un plan orthogonal.

Exercice 9. Plante associé Maïs ensilage (contexte réel, par B. Delsert)

Dans le cadre de cet exercice, on étudie l'impact d'une association culturale d'une culture de maïs ensilage. Le but est donc d'étudier l'impact de la présence ou non d'une plante compagne sur le rendement total (en tonne de matière sèche/ha/an). L'exercice s'appuie sur la problématique étudiée par le projet reine adélaïde.
Dans notre cas, on étudiera l'association Maïs en fonction de plantes compagnes Féverole, pois fourrager, Lablab (un haricot tropical), trèfle. L'association se fera donc entre le maïs et 1,2,3 ou 4 plantes compagnes, nous étudierons les diverses possibilités.

	Modalités	Semis de la plante associée	Densité semis en grains /m²	Concurrence sur le maïs	Maturité de la plante associée à la récolte	% MS du mélange	Rdt du mélange (tMS/ha)	Dont rdt plante associée (tMS/ha)	% MAT
2013	Maïs + Féverole	En même temps que le maïs	11 + 22	Très forte	Plantes sèches mais gousses pleines	33.2	10.4	4	10.6
	Maïs + Vesce + Pois fourrager		11+12+18	Moyenne	Plantes sèches et versées	27.2	9.1	0.6	7.3
	Maïs + Soja		11 + 35	Faible	Absentes	27.5	17.1	0.2	7.5
	Maïs pur	1er mai	11	/	/	27.2	18.1	/	7.6
2014	Maïs + Féverole	6 semaines après le maïs	11 + 22	Faible	Absentes	/	/	/	/
	Maïs + Haricot tarbais		11 + 5.5	Faible	Plantes vertes	29	17.6	0.4	7.9
	Maïs + Colza fourrager		11 + 5 kg/ha	Faible	Plantes vertes mais peu développées	/	/	/	/
	Maïs + Trèfle d'Alexandrie		11 + 12kg/ha	Faible	Plantes vertes mais peu développées	/	/	/	/
	Maïs pur	18 mai	11	/	/	27	17.1	/	7.6
2015	Maïs + Féverole	7 jours après le maïs	9.6 + 22	Faible (mais salissement fort impactant)	Absentes	/	/	/	/
	Maïs + Haricot tarbais		9.6 + 4.5		Plantes vertes	32.2	14.1	0.6	7.4
	Maïs + Soja		9.6 + 35		Absentes	/	/	/	/
	Maïs + Lupin		9.6 + 30		Absentes	/	/	/	/
	Maïs + Trèfles violet, Alexandrie, incarnat		9.6 + 14 kg/ha		Détruites par binage car salissement fort	/	/	/	/
	Maïs pur	13 mai	9.6	/	/	30.2	17	/	8

intéressant intermédiaire décevant

1. **Combien d'essai sont nécessaires pour construire le plan complet ?**

Dans notre cas, on étudie 4 facteurs à 2 niveaux de modalités (présence ou absence de la plante compagne). Le plan complet est donc un plan a 2^4=16 essais.

2. **On ne dispose que de 8 parcelles expérimentales. Quel plan de base choisissez-vous d'utiliser ? Construire la matrice des effets et des confusions puis construire le plan d'expérience sur R et interpréter.**

Nous avons 8 parcelles représentant 8 essais il faudra donc prendre un plan en 2^{4-1}=8essais. Nous avons donc un plan fractionnaire en résolution IV ou les effets principaux sont confondu avec les interactions d'ordre 3 ou plus.

La matrice des effets est la suivante : on observe la confusion des effets principaux avec les interactions d'ordre 3.

I	A	B	C	AB	AC	BC	ABC
ABCD	BCD	ACD	ABD	CD	BD	AD	D
1	1	1	1	1	1	1	1
1	1	1	-1	1	-1	-1	-1
1	1	-1	1	-1	1	-1	1
1	1	-1	-1	-1	-1	1	1
1	-1	1	1	-1	-1	1	-1
1	-1	1	-1	-1	1	-1	1
1	-1	-1	1	1	-1	-1	1
1	-1	-1	-1	1	1	1	-1

Le générateur d'alias est donc I=ABCD et nous avons les effets primaires confondu avec les interactions d'ordre 3 tel que :

A = BCD B=ACD C=ABD D=ABC

Sous R, il est possible de retrouver la matrice des 8 essais avec le package FrF2.

```
plan <- FrF2(nruns=8, factors=4,factor.names=list(féverole=c("feverole",""),pois=c("pois",""),la
blal=c("lablal",""),trèfle=c("trèfle","")))
plan

  féverole pois lablal trèfle
1               lablal trèfle
2
3 feverole pois
4 feverole pois lablal trèfle
5          pois        trèfle
```

```
6 feverole       lablal
7           pois lablal
8 feverole              trèfle
class=design, type= FrF2
```

D'un point de vue biologique, la confusion des effets principaux avec les interactions d'ordre 3 est acceptable. En effet les 4 plantes étudiées étant des légumineuses (capable de fixer l'azote atmosphérique), on peut penser que ces plantes jouent le même rôle, l'interaction entre les espèces ne pourrait se faire qu'au niveau de l'occupation de différentes niches (ex le trèfle au ras du sol et le Lablab grimpant sur les tiges).

Partie 2 : densité de semis

Suite à l'exercice précédent on s'intéresse ici à des densités de semis, l'idée est de définir la densité optimale de semis pour le Maïs et pour sa plante compagne afin de maximiser le rendement.

3. **On souhaite réaliser un plan composite centré afin de trouver cet optimum, décrire le modèle utilisé.**

Nous avons ici une variable quantitative expliqué par des variables quantitatives, nous allons étudier les effets linéaires et quadratiques de ces facteurs selon le modèle suivant :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon$$

Avec Y le rendement par ha, X_1 X_2 les 2 facteurs étudiées (densité de semis des 2 plantes) avec leur effets linéaires, quadratiques et d'interaction, β les constantes du modèle et ε la résiduelle suivant une loi normale d'espérance 0 et de variance 1.

4. **Combien d'essais seront nécessaire pour réaliser cette expérience ?**

Le nombre d'essai d'un plan composite centré se défini selon la formule : $2^k + 2k + n_0$ car l'on se base sur un plan factoriel complet.
Avec k le nombre de facteurs étudiés (ici 2) et n0 le nombre de points au centre (ici 8).
Le nombre d'expérience est donc $2^2+2\text{x}2+8=16$ essais.

5. **Construire le plan sur R. On étudiera l'association maïs féverole avec des densités de semis comprise entre 8 et 14 grains au m² pour le maïs et de 17 à 27 grains au m² pour la féverole.**

Nous utilisons le package rsm afin de construire le plan composite centré avec les lignes de code suivantes :

```
library(rsm)
plan <- ccd(2, coding=list (x1~(Maïs-11)/2, x2~(Feverole-22)/2))
```

Ici nous avons construit un plan composite centré avec 2 facteurs centré autour de 11 avec une étendue de 3 pour le maïs et centré autour de 22 avec une étendue de 5 pour la féverole.

```
   run.order std.order      Maïs Feverole Block
1          1         6 11.000000 22.00000     1
2          2         2 14.000000 17.00000     1
3          3         7 11.000000 22.00000     1
4          4         3  8.000000 27.00000     1
5          5         5 11.000000 22.00000     1
6          6         1  8.000000 17.00000     1
7          7         4 14.000000 27.00000     1
8          8         8 11.000000 22.00000     1
9          1         8 11.000000 22.00000     2
10         2         1  6.757359 22.00000     2
11         3         5 11.000000 22.00000     2
12         4         6 11.000000 22.00000     2
13         5         2 15.242641 22.00000     2
14         6         3 11.000000 14.92893     2
15         7         4 11.000000 29.07107     2
16         8         7 11.000000 22.00000     2

Data are stored in coded form using these coding formulas ...
x1 ~ (Maïs - 11)/3
x2 ~ (Feverole - 22)/5
```

6. **Etudier la qualité du plan**

Pour étudier la qualité du plan, on calcule la matrice des effets et on étudie $(X'X)^{-1}$.

```
X <- model.matrix(~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data=plan)
```

```
solve(t(X)%*%X)

              (Intercept)    x1    x2 I(x1^2) I(x2^2) I(x1 * x2)
(Intercept)        0.1250 0.000 0.000 -0.0625 -0.0625       0.00
x1                 0.0000 0.125 0.000  0.0000  0.0000       0.00
x2                 0.0000 0.000 0.125  0.0000  0.0000       0.00
I(x1^2)           -0.0625 0.000 0.000  0.1250  0.0000       0.00
I(x2^2)           -0.0625 0.000 0.000  0.0000  0.1250       0.00
I(x1 * x2)         0.0000 0.000 0.000  0.0000  0.0000       0.25
```

Nous avons ici une matrice qui est « presque » parfaite car la constante est indépendante des effets principaux (présence de 0 au niveau des lignes x1 et x2 dans la colonne intercept). De plus, les effets principaux et le produits x1*x2 sont indépendants de tous les autres effets et les effets quadratiques sont faiblement lié à la constante (-0.0625 proche de 0). Nous avons donc une matrice qui montre que les essais sont assez bien choisis :la qualité du plan est donc optimale étant donné les contraintes imposées par le contexte.

7. **Un nouvel expérimentateur souhaite réitérer cette expérience mais doit répartir ses essais dans 2 parcelles, décrire la démarche que vous lui conseillez d'adopter et comment analyser les résultats.**

Dans le contexte qui nous ai proposé, on réalise les essais sur une même parcelle, dans ce nouveau cas, il faudrait prendre en compte un effet parcelle qui correspondrai à un nouveau facteur à étudier. Pour cela l'expérimentateur devra s'orienter vers des plans optimaux, on peut proposer l'utilisation du package AlgDesign qui, via la fonction optfederov va donner les essais à réaliser pour ce type de contexte.

```
Mais<-seq(8,14,by=1)
Feverole<-seq(17,27,by=1)
Parcelle<-seq(1,2,by=1)
dat<-expand.grid(list(A=Mais,B=Feverole,C=Parcelle))
set.seed(1234)
plan <- optFederov(~A+B+I(A^2)+I(B^2)+I(A*B)+C, data=dat, nTrials=16 )
```

Ici nous avons une grille regroupant toutes les combinaisons possibles, le nombre d'essais étant limité, on obtient les essais suivants :

```
$design
     A  B C
1    8 17 1
4   11 17 1
7   14 17 1
39  11 22 1
42  14 22 1
71   8 27 1
77  14 27 1
78   8 17 2
81  11 17 2
84  14 17 2
113  8 22 2
116 11 22 2
119 14 22 2
148  8 27 2
151 11 27 2
154 14 27 2
```

La première colonne étant la ligne à laquelle on retrouve la combinaison dans la grille complète (dat), les colonnes A et B représentent les densités de semis de Maïs et Féverole et le Facteur C représentant les 2 parcelles.

```
X <- model.matrix(~A+B+C+I(A^2)+I(B^2)+I(A*B),data=plan$design)
round(solve(t(X)%*%X),2)

            (Intercept)      A      B     C I(A^2) I(B^2) I(A * B)
(Intercept)      168.25 -10.78 -10.31  0.04   0.35   0.20     0.13
A                -10.78   2.10  -0.05 -0.03  -0.08   0.00    -0.01
B                -10.31  -0.05   1.00 -0.03   0.01  -0.02    -0.01
C                  0.04  -0.03  -0.03  0.26   0.00   0.00     0.00
I(A^2)             0.35  -0.08   0.01  0.00   0.00   0.00     0.00
I(B^2)             0.20   0.00  -0.02  0.00   0.00   0.00     0.00
I(A * B)           0.13  -0.01  -0.01  0.00   0.00   0.00     0.00
```

Le plan est donc un peu détérioré par rapport au précédent qui ne tenait pas compte du facteur C.

Exercice 10. Brassage d'une bière (exemple fictif, par Laurine Allard)

L'exercice proposé est composé de deux parties : une première partie concernant le brassage d'une bière et la seconde, la dégustation d'une bière. La thématique et les données sont fictives.

Partie 1 : Realisation d'un brassage

Un brasseur amateur souhaite réaliser une bière de style IPA, avec un taux d'amertume fixé entre 40 et 60 IBU (IBU est l'indice mesurant le taux d'amertume dans une bière). Pour trouver la recette idéale à partir des ingrédients qu'il dispose, le brasseur souhaite réaliser plusieurs essais en faisant varier les quantités de chacun. Deux ingrédients sont susceptibles d'influencer le taux d'amertume : le houblon et le malt. Ainsi, ces deux facteurs seront étudiés et des brassins de 10 litres seront réalisés :

- Houblon : la quantité doit être comprise entre 40 et 110 grammes.
- Malt : la quantité doit être comprise entre 2 et 3.2 kilogrammes.

Pour des questions de coût et de temps, il souhaite réaliser le moins d'essais possibles.

1. **Construire le plan d'expériences sur R.**

Etant donné que nous étudions des variables quantitatives, il faut réaliser un plan composite centré. Voici les lignes de code R pour la construction du plan d'expériences.

```
library(rsm)
plan <- ccd(2, n0=2, randomize=FALSE, coding=list (x1~(houblon-75)/35, x2~(malt-2.6)/0.6))
plan
   run.order std.order   houblon     malt Block
1          1         1  40.00000 2.000000     1
2          2         2 110.00000 2.000000     1
3          3         3  40.00000 3.200000     1
4          4         4 110.00000 3.200000     1
5          5         5  75.00000 2.600000     1
6          6         6  75.00000 2.600000     1
7          1         1  25.50253 2.600000     2
8          2         2 124.49747 2.600000     2
9          3         3  75.00000 1.751472     2
10         4         4  75.00000 3.448528     2
11         5         5  75.00000 2.600000     2
12         6         6  75.00000 2.600000     2
```

12 essais sont donc à réaliser (points au centre, points en étoile et les points du plan factoriel complet).

2. **Quel est le modèle à utiliser pour analyser les résultats d'un tel plan ?**

Le modèle comporte, pour chaque variable explicative X1 (houblon) et X2 (malt), un effet linéaire et un effet quadratique, auxquels s'ajoute un terme d'interaction entre X1 et X2 :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X^2{}_1 + \beta_{22} X^2{}_2 + \beta_{12} X_1 X_2 + \varepsilon$$

À ce modèle, s'ajoutent les hypothèses usuelles sur les résidus ε.

3. **A la suite de la réalisation de ces essais, le brasseur a mesuré le taux d'amertume (IBU) pour chacun des brassins. Voici les résultats qu'il a obtenu : 45, 58, 51, 62, 50, 50, 43, 50, 48, 57, 50, 50. A l'aide de la fonction rsm du package rsm, analyser les résultats. Quelle est l'erreur pure ? Que pouvez-vous dire sur l'ajustement du modèle ? Quel modèle conservez-vous ?**

Voici les lignes de code R :

```
Y <- c(45, 58, 51, 62, 50, 50, 43, 50, 48, 57, 50, 50)
CR.rsm <- rsm(Y~SO(x1,x2),data=plan)
summary(CR.rsm)
Call:
rsm(formula = Y ~ SO(x1, x2), data = plan)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  50.0000     1.6619 30.0866 8.944e-08 ***
x1            4.2374     1.1751  3.6060   0.01129 *
x2            2.8410     1.1751  2.4176   0.05203 .
```

```
x1:x2        -0.5000     1.6619 -0.3009   0.77367
x1^2         -0.6250     1.3138 -0.4757   0.65111
x2^2          2.3750     1.3138  1.8077   0.12066
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.7926,     Adjusted R-squared:  0.6199
F-statistic: 4.587 on 5 and 6 DF,  p-value: 0.04544

Analysis of Variance Table

Response: Y
            Df  Sum Sq Mean Sq    F value  Pr(>F)
FO(x1, x2)   2 208.217 104.108 9.4240e+00 0.01408
TWI(x1, x2)  1   1.000   1.000 9.0500e-02 0.77367
PQ(x1, x2)   2  44.167  22.083 1.9990e+00 0.21613
Residuals    6  66.283  11.047
Lack of fit  3  66.283  22.094 1.0156e+31 < 2e-16
Pure error   3   0.000   0.000

Stationary point of response surface:
        x1         x2
 3.4825568 -0.2315183

Stationary point in original units:
   houblon       malt
196.889488   2.461089

Eigenanalysis:
eigen() decomposition
$values
[1]  2.3956906 -0.6456906

$vectors
          [,1]        [,2]
x1 -0.08248053 -0.99659268
x2  0.99659268 -0.08248053
```

Nous observons que l'erreur d'ajustement est inférieure à 5% (< 2e-16) donc elle est significative. L'erreur pure quant à elle est égale à 0. Ce modèle ne semble pas bien ajusté ($R^2 = 0.7926$). Seuls les coefficients X1 et X2 sont significatifs.

4. **Ecrire le modèle finalement conservé pour la recherche de l'optimum et donner la valeur du couple (houblon, malt) pour lequel le taux d'amertume est optimum ?**

Seuls les coefficients linéaires sont significatifs. On retient donc le modèle sans interaction et sans effet quadratique. Voici les lignes code R pour le nouveau modèle :

```
Y <- c(45, 58, 51, 62, 50, 50, 43, 50, 48, 57, 50, 50)
CR.rsm <- rsm(Y~FO(x1,x2),data=plan)
summary(CR.rsm)

Call:
rsm(formula = Y ~ FO(x1, x2), data = plan)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  51.1667     1.0158 50.3685 2.405e-12 ***
x1            4.2374     1.2442  3.4059   0.00780 **
x2            2.8410     1.2442  2.2835   0.04829 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.6514,     Adjusted R-squared:  0.5739
F-statistic: 8.407 on 2 and 9 DF,  p-value: 0.008724

Analysis of Variance Table

Response: Y
            Df Sum Sq Mean Sq    F value    Pr(>F)
FO(x1, x2)   2 208.22 104.108 8.4071e+00  0.008724
Residuals    9 111.45  12.383
Lack of fit  6 111.45  18.575 8.5381e+30 < 2.2e-16
Pure error   3   0.00   0.000

Direction of steepest ascent (at radius 1):
       x1        x2
0.8305970 0.5568739

Corresponding increment in original units:
   houblon       malt
```

```
29.0708962   0.3341244
```

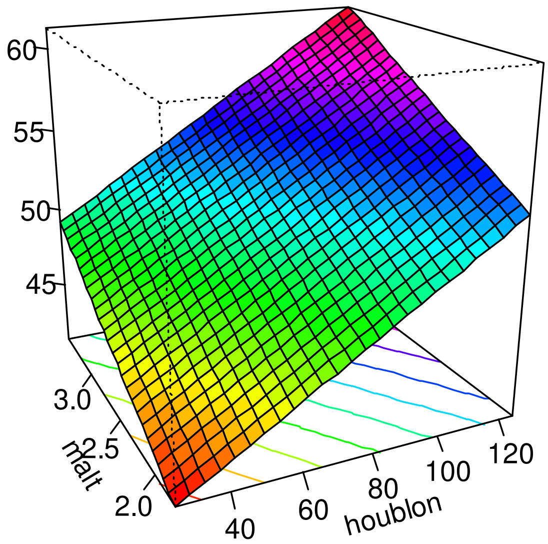
Ainsi, $Y = 51{,}1667 + 4.2374\ X_1 + 2.8410\ X_2$ et $X_1 = 0.83$; $X_2 = 0.56$

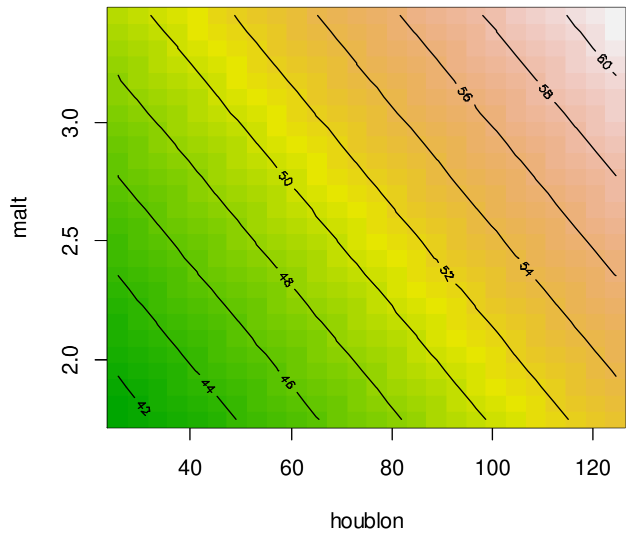
Donc $Y = 51.1667 + 4.2374 * 0.83 + 2.8410 * 0.56 = 56$ IBU

Ainsi le couple (houblon, malt) pour lequel le taux d'amertume est optimum avec 56 IBU est (29 grammes ; 0.33 kg).

5. **A l'aide des fonctions contour ou persp, représenter les surfaces de réponses et dire si l'optimum est un minimum ou un maximum ?**

```
contour(CR.rsm,~x1+x2,image=TRUE)
persp(CR.rsm,~x1+x2,contours="colors")
```





L'optimum est un minimum.

Partie 2 : Degustation d'une biere

Après sélection de la meilleure recette de bière par le brasseur amateur, celui-ci se demande s'il existe un effet de l'environnement de dégustation sur la perception de l'amertume de sa bière IPA pour un individu i. Pour cela, il souhaite réaliser des dégustations à une personne de son entourage selon différents facteurs. Cette personne dégustera la même bière sans le savoir à chaque essai. La température de service sera identique pour chaque essai (3°C). Le brasseur souhaite prendre en compte les facteurs suivants dans ses dégustations :

- **Facteur fond sonore** : dégustation en silence, lors d'une discussion, fond musical.
- **Facteur luminosité** : dégustation à la lumière du jour, dégustation en aveugle, dégustation sous une lumière rouge (box d'analyse sensorielle).
- **Facteur moment de dégustation** : matin, midi, soir.
- **Facteur météorologique** : temps ensoleillé, temps nuageux, temps pluvieux.

Pour chaque dégustation, le panéliste évaluera l'amertume de la bière sur une échelle d'intensité de 0 (pas du tout amer) à 10 (extrêmement amer). Le brasseur souhaite réaliser au maximum 9 dégustations.

6. **Construire le plan d'expériences et décrire les essais à réaliser.**

Etant donné qu'on étudie 4 facteurs à 3 modalités, nous pouvons construire un MOLS (carrés latins mutuellement orthogonaux).

		F2		
		0	1	2
F1	0	0	1	2
	1	1	2	0
	2	2	0	1

F3

		F2		
		0	1	2
F1	0	0	1	2
	1	2	0	1
	2	1	2	0

F4

Essais	Fond sonore	Luminosité	Moment de dégustation	Météo
E1	Silence	Lumière du jour	Matin	Ensoleillé
E2	Silence	Aveugle	Midi	Nuageux
E3	Silence	Lumière rouge	Soir	Pluvieux
E4	Discussion	Lumière du jour	Midi	Pluvieux
E5	Discussion	Aveugle	Soir	Ensoleillé
E6	Discussion	Lumière rouge	Matin	Nuageux
E7	Musical	Lumière du jour	Soir	Nuageux
E8	Musical	Aveugle	Matin	Pluvieux
E9	Musical	Lumière rouge	Midi	Ensoleillé

7. **Quelles sont les propriétés de ce plan ?**

Ici, 4 facteurs de 3 modalités sont étudiés. Les modalités sont donc équilibrées. Par construction, les effets principaux sont orthogonaux deux à deux.

8. **Quel sera le modèle utilisé pour analyser les résultats ? Préciser les degrés de liberté associés aux paramètres du modèle.**

Le modèle utilisé pour analyser les résultats sera une analyse de variance à 4 facteurs comprenant les effets principaux.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl}$$

2 ddl (3 – 1) sont utilisés par facteur et 1 ddl pour la constante, ce qui fait donc 9 ddl.

9. **Est-il possible d'introduire d'autres facteurs d'environnement de dégustation, comme par exemple le fait que le dégustateur ait consommé quelque chose dans l'heure précédent la dégustation de la bière (aliment sucré, aliment salé, rien) ?**

Etant donné que le brasseur souhaitait réaliser 9 essais, il ne pourra pas étudier de facteurs supplémentaires car le nombre de degrés de liberté est limité.

10. **A la suite de la réalisation de ces dégustations, le brasseur a obtenu les notes d'intensité d'amertume selon les essais décrits précédemment. Les voici :**

Essais	Notes
E1	**8**
E2	**5**
E3	**6**
E4	**7**
E5	**7**
E6	**4**
E7	**3**
E8	**8**
E9	**5**

Réaliser l'analyse des résultats sur R.

Après enregistrement des données sur un fichier Excel, voici les lignes de commande à rentrer sur R.

```
library(FactoMineR)
amertume <- read.table("amertume.csv", header = TRUE, sep = ";")
summary(amertume)
df   <-   data.frame(fond=unlist(fond),   luminosite=unlist(luminosite),   moment=unlist(moment),
meteo=unlist(meteo), note=unlist(note))
options(contras=c("contr.sum","contr.sum"))
amertume.anova1 <- aov(note ~ fond, data = df)
summary(amertume.anova1)
            Df Sum Sq Mean Sq F value Pr(>F)
fond         2      0       0       0      1
Residuals    6     18       3
amertume.anova2 <- aov(note ~ luminosite, data = df)
```

```
summary(amertume.anova2)
            Df Sum Sq Mean Sq F value Pr(>F)  
luminosite   2 12.667   6.333   7.125  0.026 *
Residuals    6  5.333   0.889                 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

amertume.anova3 <- aov(note ~ moment, data = df)
summary(amertume.anova3)
            Df Sum Sq Mean Sq F value Pr(>F)
moment       2  0.667  0.3333   0.115  0.893
Residuals    6 17.333  2.8889               

amertume.anova4 <- aov(note ~ meteo, data = df)
summary(amertume.anova4)
            Df Sum Sq Mean Sq F value Pr(>F)
meteo        2  4.667   2.333    1.05  0.406
Residuals    6 13.333   2.222               
```

11. Que pouvez-vous conclure sur la perception de l'amertume pour un individu selon son environnement de dégustation ?

D'après les résultats obtenus sur R, il n'y a que le facteur luminosité qui a un effet sur la perception de l'amertume car c'est le seul facteur qui a une p-value significative (<5%).

Exercice 11. Analyse sensorielle de punchs (exemple fictif, par C. Danet)

Partie 1 : Analyse sensorielle d'un ensemble de punchs

Dans le cadre de l'organisation d'un évènement très important, une entreprise issue de l'industrie agroalimentaire souhaite préparer le punch « idéal » pour ses clients. De nombreuses recettes existent et elle souhaiterait déterminer la meilleure. Pour cela, l'entreprise souhaite réaliser une analyse sensorielle sur un lot de punchs présentant une composition légèrement différente. Certains éléments comme le jus d'orange et le jus d'ananas resteront les mêmes d'un punch à l'autre, mais d'autres varieront. L'entreprise souhaite ainsi jouer sur les 7 facteurs suivants, avec deux possibilités pour chacun :

- Oranges : présence de morceaux / absence
- Bananes : du jus / des fruits
- Rhum : blanc / ambré
- Sucre de canne : sous forme de sirop / sous forme de grains
- Gousse de vanille : présente / absente
- Bâton de cannelle : présent / absent
- Jus exotique : jus de mangue / jus de goyave

Cependant, le délai avant l'évènement est court et il n'est pas raisonnable de faire tester aux sujets plus de 8 punchs différents. L'objectif ici est donc de décrire précisément les 8 punchs que l'entreprise devra faire goûter aux sujets lors de l'analyse sensorielle.

1. Décrire le plan à réaliser.

Le plan à réaliser comprend 7 variables à 2 modalités. Il est à réaliser en 8 essais. Il faut donc construire un plan 2^{7-4}.

2. Construire le plan sous R, donner les générateurs d'alias et la résolution du plan.

Pour étudier les 7 facteurs en 8 essais, on choisit d'abord le plan de base, puis les interactions avec lesquelles on confond les facteurs qu'on ajoute, et enfin on liste les confusions engendrées.
La matrice des effets correspondant au plan 2^3 présente les colonnes suivantes :
I A B C AB AC BC ABC
4 facteurs doivent être ajoutés à ce plan de base pour avoir les 7 effets attendus. On va donc confondre un facteur D avec l'interaction AB, un E avec AC, F avec BC et G avec ABC.

On obtient alors les générateurs d'aliases suivants : I_1=ABD, I_2=ACE, I_3=BCF, I_4=ABCG.
En réalisant les produits de ces générateurs, on obtient les autres générateurs : I_{12} = BCDE ; I_{13} = ACDF ; I_{14} = CDG ; I_{23} = ABEF ; I_{24} = BEG ; I_{34} = AFG ; I_{123} = DEF ; I_{124} = ADEG ; I_{134} = BDFG ; I_{234} = CEFG ; I_{1234} = ABCDEFG. On a bien ici les 15 générateurs attendus.
La plus petite interaction avec laquelle la constante est confondue est d'ordre 3, le plan est donc de résolution III. On émet donc l'hypothèse que les interactions d'ordre 2 et plus sont négligeables, de manière à pouvoir estimer les effets principaux.

Le script effectué sur R et les résultats obtenus sont les suivants :

```
library(FrF2)
plan_punchs <- FrF2(nruns=8,nfactors=7,factor.names=list(Morceaux_oranges=c("présence","absence"
), Bananes=c("jus","fruits"), Rhum=c("blanc","ambré"), Sucre=c("sirop","grains"), Vanille=c("pré
sence","absence"), Cannelle=c("présence","absence"), Jus_exotique=c("mangue","goyave"))) #Constr
uction du plan des punchs à déguster
summary(plan_punchs) #Obtiention du plan à présenter directement aux dégustateurs

Call:
FrF2(nruns = 8, nfactors = 7, factor.names = list(Morceaux_oranges = c("présence",
    "absence"), Bananes = c("jus", "fruits"),
    Rhum = c("blanc", "ambré"), Sucre = c("sirop",
        "grains"), Vanille = c("présence", "absence"),
    Cannelle = c("présence", "absence"), Jus_exotique = c("mangue",
        "goyave")))

Experimental design of type  FrF2
8  runs

Factor settings (scale ends):
  Morceaux_oranges Bananes  Rhum  Sucre  Vanille Cannelle Jus_exotique
1         présence     jus blanc  sirop présence présence       mangue
2          absence  fruits ambré grains  absence  absence       goyave

Design generating information:
$legend
[1] A=Morceaux_oranges B=Bananes          C=Rhum              D=Sucre             E=Vanille

[6] F=Cannelle         G=Jus_exotique

$generators
[1] D=AB  E=AC  F=BC  G=ABC


Alias structure:
$main
[1] A=BD=CE=FG B=AD=CF=EG C=AE=BF=DG D=AB=CG=EF E=AC=BG=DF F=AG=BC=DE G=AF=BE=CD


The design itself:
  Morceaux_oranges Bananes  Rhum  Sucre  Vanille Cannelle Jus_exotique
1         présence  fruits ambré  sirop présence  absence       mangue
2         présence     jus blanc grains  absence  absence       mangue
3          absence  fruits ambré grains  absence  absence       goyave
4          absence     jus blanc  sirop présence  absence       goyave
5         présence  fruits blanc  sirop  absence présence       goyave
6         présence     jus ambré grains présence présence       goyave
7          absence     jus ambré  sirop  absence présence       mangue
8          absence  fruits blanc grains présence présence       mangue
class=design, type= FrF2
```

3. Analyser la qualité du plan obtenu.

Pour analyser la qualité du plan, le script R effectué et le tableau obtenu sont les suivants :

```
X <- model.matrix(~.,data=plan_punchs) #Construction de la matrice des confusions
solve(t(X)%*%X)
                 (Intercept) Morceaux_oranges1 Bananes1 Rhum1 Sucre1 Vanille1 Cannelle1 Jus_exotique1
(Intercept)            0.125             0.000    0.000 0.000  0.000    0.000     0.000         0.000
Morceaux_oranges1      0.000             0.125    0.000 0.000  0.000    0.000     0.000         0.000
Bananes1               0.000             0.000    0.125 0.000  0.000    0.000     0.000         0.000
Rhum1                  0.000             0.000    0.000 0.125  0.000    0.000     0.000         0.000
Sucre1                 0.000             0.000    0.000 0.000  0.125    0.000     0.000         0.000
Vanille1               0.000             0.000    0.000 0.000  0.000    0.125     0.000         0.000
Cannelle1              0.000             0.000    0.000 0.000  0.000    0.000     0.125         0.000
Jus_exotique1          0.000             0.000    0.000 0.000  0.000    0.000     0.000         0.125
```

Ce plan est de bonne qualité, il n'y a pas d'ambiguïté entre les effets principaux puisqu'on obtient pour la matrice (X'X)-1 une matrice diagonale avec un coefficient de 0,125. Les covariances entre les effets sont nulles.

4. **Emettre des hypothèses concernant les éventuelles interactions entre les facteurs de composition des punchs. Donner des pistes pour analyser ces interactions.**

On peut suspecter une interaction entre la vanille et la cannelle, deux condiments forts, pouvant s'annuler mutuellement si on les ajoute ensemble dans le punch.
Cependant, les 7 facteurs nécessitent l'estimation de 8 paramètres (les 7 variables et la constante). Le plan étant en 8 essais, il est donc saturé. Il n'est donc pas possible d'estimer plus d'effets que ceux des 7 facteurs ainsi que la constante.
De manière à prendre en compte cette interaction, il faudrait construire un plan avec plus d'essais, en en ajoutant au moins 2.

PARTIE 2 – PLANS DE DEGUSTATION DES PUNCHS

La même entreprise veut désormais faire déguster les 8 punchs choisis par 8 juges. Chacun des juges doit déguster chacun des punchs une fois. Ils auront pour consigne de noter de 0 à 10 chaque punch selon 6 critères : goût sucré, goût fruité, équilibre rhum/fruits, acidité, goût exotique, odeur de rhum. Ils devront également donner une note finale d'appréciation pour chaque produit. Les punchs seront présentés de façon monadique séquentielle (c'est-à-dire un par un). L'entreprise souhaite éviter toute potentielle influence du rang de dégustation sur les notations des différents produits. En prenant en compte ces contraintes, vous devez établir le plan de dégustation des 8 punchs en indiquant pour chaque juge l'ordre dans lequel il devra gouter les produits.

5. **Construire le plan en expliquant vos choix.**

Pour la dégustation on retrouve 3 facteurs : le juge, le punch et le rang de dégustation. Il y a 8 juges, 8 punchs et 8 rangs de dégustation. On peut donc construire un carré latin.
Pour la première ligne, j'ai rangé les punchs du premier au huitième. Puis, j'ai obtenu les 7 lignes suivantes par permutations circulaires. Le plan obtenu est le suivant :

	Rang 1	Rang 2	Rang 3	Rang 4	Rang 5	Rang 6	Rang 7	Rang 8
Juge 1	1	2	3	4	5	6	7	8
Juge 2	2	3	4	5	6	7	8	1
Juge 3	3	4	5	6	7	8	1	2
Juge 4	4	5	6	7	8	1	2	3
Juge 5	5	6	7	8	1	2	3	4
Juge 6	6	7	8	1	2	3	4	5
Juge 7	7	8	1	2	3	4	5	6
Juge 8	8	1	2	3	4	5	6	7

6. **Donner les propriétés du plan.**

Chaque juge évalue tous les punchs, les facteurs punch et juge sont donc orthogonaux. Cependant, si tous les punchs sont dégustés le même nombre de fois à chaque rang, on retrouve tout de même une confusion partielle pour les interactions punch:rang et juge:rang.

7. **Ecrire le modèle qui sera utilisé pour analyser les résultats obtenus, en détaillant les termes et les degrés de liberté associés.**

Avec un nombre égal de juges et de punchs, le modèle ne peut donc pas contenir les effets des interactions, mais uniquement les effets principaux. Il s'écrira donc :
$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$;
Avec : μ la constante associée à l'effet moyen général
α_i le paramètre associé à l'effet du punch i
β_j le paramètre associé à l'effet du juge j
γ_k le paramètre associé à l'effet du rang k
ε_{ijk} l'erreur résiduelle du modèle
Pour chaque facteur, 7 degrés de liberté (ddl) sont utilisés. 1 ddl est utilisé pour la constante.
Il reste donc (8*8) – (7*3 + 1) = 42 ddl pour la résiduelle.

8. **Critiquer le modèle obtenu en analysant les interactions non prises en compte.**

Comme dit précédemment, l'interaction punch:rang génère une confusion non prise en compte dans le modèle. On pourrait alors craindre que l'ordre de succession des produits, toujours le même ici, ait une influence sur la notation des juges. Un nouveau plan pourrait donc être envisagé, dans lequel chaque punch devrait être dégusté le même nombre de fois à la suite de chacun des autres, pour éviter l'influence de l'ordre de dégustation sur le ressenti des juges.

Exercice 12. Mise en place d'un plan d'expérience pour des émulsions (exemple réel, Pauline HENRIET)

Lors de mon stage M1 en laboratoire de recherche, j'ai cherché à réaliser des émulsions (huile dans l'eau) stabilisées par de la gélatine de peau de poisson pour les caractériser et trouver leurs domaines d'application. En effet, la gélatine est connue de façon générale pour ses propriétés stabilisantes et émulsifiantes.

PARTIE 1 : PLANS FRACTIONNAIRES ET ASYMETRIQUES

On cherche dans un premier temps à former des émulsions avec les plus petites gouttelettes, qui sont supposées être les plus stables dans le temps. Pour cela, on évalue l'impact des traitements mécaniques appliqués pour former des émulsions. Deux traitements mécaniques sont choisis pour former ces émulsions :

- *La dispersion mécanique (facteur A) : à 8000 rpm ou 13500 rpm*
- *Les ultrasons (facteur B) : pendant 20 ou 20 minutes*

On peut également faire varier :

- *La concentration en huile (facteur C) : 2% ou 5%*
- *La concentration en gélatine dans la phase aqueuse (facteur D) : 2% ou 4%*
- *Le mode d'incorporation de l'huile (facteur E) : en une seule fois ou goutte par goutte.*

Une phase expérimentale est lancée, on estime que le traitement mécanique le plus efficace est celui nous permettant d'obtenir les plus petits globules d'huiles (réponse Y).

1. **Indiquer le nombre d'essais à réaliser pour ne pas avoir de confusions, et donner sa résolution**

Le nombre d'essais à réaliser pour ne pas avoir de confusion est de : 2^5 = 32 essais. On est alors avec une résolution IV

2. **Cependant les matières premières étant coûteuses et en quantité limitée, seulement 8 essais sont possibles. Construire le plan sur R et déduire les générateurs d'alias et la résolution du plan obtenu.**

```
library (FrF2)

plexp<-FrF2(nruns=8, nfactors=5, factor.names = list(mecanique=c("8000rpm", "13500rpm"), ultrasons=c("20min","30min"), huile=c("2%", "5%"), gelatine=c("2%", "4%"), incorporation=c("une fois", "goutte/goutte")))

summary(plexp2)
```

```
Call:
FrF2(nruns = 8, nfactors = 5, factor.names = list(mecanique = c("8000rpm",
    "13500rpm"), ultrasons = c("20min", "30min"), huile = c("2%",
    "5%"), gelatine = c("2%", "4%"), incorporation = c("une fois",
    "goutte/goutte")))

Experimental design of type  FrF2
8  runs

Factor settings (scale ends):

Design generating information:
$legend
[1] A=mecanique     B=ultrasons     C=huile         D=gelatine      E=incorporation

$generators
[1] D=AB  E=AC
```

```
Alias structure:
$main
[1] A=BD=CE B=AD     C=AE     D=AB     E=AC

$fi2
[1] BC=DE BE=CD
```

	mecanique <fctr>	ultrasons <fctr>	huile <fctr>	gelatine <fctr>	incorporation <fctr>
1	8000rpm	20min	2%	4%	goutte/goutte
2	8000rpm	30min	2%	2%	goutte/goutte
3	13500rpm	30min	5%	4%	goutte/goutte
4	13500rpm	30min	2%	4%	une fois
5	13500rpm	20min	2%	2%	une fois
6	8000rpm	20min	5%	4%	une fois
7	13500rpm	20min	5%	2%	goutte/goutte
8	8000rpm	30min	5%	2%	une fois

On passe donc en résolution 3 : les effets principaux sont confondus avec des interactions d'ordre 2. Les générateurs d'alias sont les suivants : D=AB E=AC

Une nouvelle stagiaire arrive lors de la phase expérimentale et va également travailler sur les émulsions, mais sur la synergie possible entre la gélatine de peau de poisson et de pois. On n'étudie plus le facteur D, mais le facteur F, ratio gélatine de poisson/protéine de pois, pouvant prendre 4 modalités (25-75, 50-50,75-25, 0-100). De plus, devant former la nouvelle stagiaire, les 8 essais sont répartis entre moi et elle (facteur G).

3. **Construire le plan $L_8 2^4 4$. Est-il de bonne qualité ? Si non, utiliser R pour l'améliorer.**

```
plan<-oa.design(nlevels =c(2,2,2,2,4), factor.names =list(mecanique=c("8000rpm", "13500rpm"),
ultrasons=c("20min","30min"), huile=c("2%", "5%"),stagiaire=c("Pauline ", "Mar"), ratio=c("25-
75", "50-50", "75-25", "0-100")))
summary(plan)
X<-model.matrix(~.,plan2)
solve(t(X)%*%X)
```

```
Factor settings (scale ends):
Generating Orthogonal Array:
[1] L8.2.4.4.1
Selected Columns:
[1] 1 2 3 4 5
Numbers of generalized words of lengths 3 and 4
          (Intercept) mecanique1 ultrasons1 huile1 stagiaire1 ratio50-50 ratio75-25 ratio0-100
(Intercept)         0.5      0.000      0.000  0.000      0.000       -0.5       -0.5       -0.5
mecanique1          0.0      0.125      0.000  0.000      0.000        0.0        0.0        0.0
ultrasons1          0.0      0.000      0.125  0.000      0.000        0.0        0.0        0.0
huile1              0.0      0.000      0.000  0.125      0.000        0.0        0.0        0.0
stagiaire1          0.0      0.000      0.000  0.000      0.125        0.0        0.0        0.0
ratio50-50         -0.5      0.000      0.000  0.000      0.000        1.0        0.5        0.5
ratio75-25         -0.5      0.000      0.000  0.000      0.000        0.5        1.0        0.5
ratio0-100         -0.5      0.000      0.000  0.000      0.000        0.5        0.5        1.0
```

	mecanique <fctr>	ultrasons <fctr>	huile <fctr>	stagiaire <fctr>	ratio <fctr>
1	13500rpm	30min	2%	Pauline	75-25
2	13500rpm	20min	5%	Pauline	50-50
3	13500rpm	20min	2%	Mar	0-100
4	8000rpm	30min	5%	Pauline	0-100
5	13500rpm	30min	5%	Mar	25-75
6	8000rpm	20min	2%	Pauline	25-75
7	8000rpm	30min	2%	Mar	50-50
8	8000rpm	20min	5%	Mar	75-25

Ici, on peut voir que le plan n'est pas orthogonal. Il y a des confusions avec : I (Intercept) et le facteur à 4 modalités (ratio de protéines).

Nouveau plan:

```
library(DoE.base)
plan2a <- oa.design(nruns= 8, nlevels=c(2,2,2,2,4))
X <- model.matrix(~.,plan2a)
round(t(X)%*%X,0)

             (Intercept) A1 B1 C1 D1 E.L E.Q E.C
(Intercept)            8  0  0  0  0   0   0   0
A1                     0  8  0  0  0   0   0   0
B1                     0  0  8  0  0   0   0   0
C1                     0  0  0  8  0   0   0   0
D1                     0  0  0  0  8   0   0   0
E.L                    0  0  0  0  0   2   0   0
E.Q                    0  0  0  0  0   0   2   0
E.C                    0  0  0  0  0   0   0   2
```

On peut dire que maintenant le plan est pseudo-orthogonal : chacune des lignes de la matrice sont indépendantes, mais les termes de la matrice ont 8 et 2 (pas égaux).

PARTIE 2 : PLANS POUR SURFACE DE REPONSES

Suite à une réunion avec le tuteur, il est décidé que le facteur B est fixé sur 20 minutes à une amplitude de 100% et que le facteur E est fixé en incorporation « goutte par goutte ». Les autres facteurs varient de la façon suivante :

- Facteur A : entre 8 000 et 20 500 rpm
- Facteur C : entre 1% et 10% d'huile
- Facteur D : entre 2% et 6% pour la concentration en protéine de poisson

On rappelle que l'objectif est d'obtenir des gouttelettes d'huile les plus petites possibles.

4. **Proposer un nouveau plan d'expérience indiquer sa qualité et le nombre d'expérience à réaliser et le nombre de points au centre avec un tel plan. Est-il possible de l'améliorer ?**

```
library(rsm)
planNouveau<-ccd(3, coding = list(x1~(dispersion-14250)/6250, x2~(huile-5.5)/4.5, x3~(gelatine-4)/2))
planNouveau
X<-model.matrix(~x1+x2+x3+I(x1^2)+I(x2^2)+I(x3^2)+I(x1*x2)+I(x2*x3)+I(x1*x3), data=planNouveau)
t(X)%*%X
run.order std.order dispersion     huile  gelatine Block
1          1        10  14250.000  5.500000 4.0000000     1
2          2         3   8000.000 10.000000 2.0000000     1
3          3         5   8000.000  1.000000 6.0000000     1
4          4         6  20500.000  1.000000 6.0000000     1
5          5        12  14250.000  5.500000 4.0000000     1
6          6         8  20500.000 10.000000 6.0000000     1
7          7         9  14250.000  5.500000 4.0000000     1
8          8        11  14250.000  5.500000 4.0000000     1
9          9         2  20500.000  1.000000 2.0000000     1
10        10         7   8000.000 10.000000 6.0000000     1
11        11         1   8000.000  1.000000 2.0000000     1
12        12         4  20500.000 10.000000 2.0000000     1
13         1         4  14250.000 13.715838 4.0000000     2
14         2         2  25660.887  5.500000 4.0000000     2
15         3         8  14250.000  5.500000 4.0000000     2
16         4         7  14250.000  5.500000 4.0000000     2
17         5         1   2839.113  5.500000 4.0000000     2
18         6         3  14250.000 -2.715838 4.0000000     2
19         7         5  14250.000  5.500000 0.3485163     2
20         8        10  14250.000  5.500000 4.0000000     2
21         9         9  14250.000  5.500000 4.0000000     2
22        10         6  14250.000  5.500000 7.6514837     2

Data are stored in coded form using these coding formulas ...
x1 ~ (dispersion - 14250)/6250
x2 ~ (huile - 5.5)/4.5
x3 ~ (gelatine - 4)/2
                  (Intercept)  x1       x2       x3       I(x1^2)  I(x2^2   I(x3^2)   I(x1 * x2 * x3)
(Intercept)          22.00000  0.00000  0.00000  0.00000 14.66667 14.66667 14.66667                0
x1                    0.00000 14.66667  0.00000  0.00000  0.00000  0.00000  0.00000                0
x2                    0.00000  0.00000 14.66667  0.00000  0.00000  0.00000  0.00000                0
x3                    0.00000  0.00000  0.00000 14.66667  0.00000  0.00000  0.00000                0
I(x1^2)              14.66667  0.00000  0.00000  0.00000 30.22222  8.00000  8.00000                0
I(x2^2)              14.66667  0.00000  0.00000  0.00000  8.00000 30.22222  8.00000                0
I(x3^2)              14.66667  0.00000  0.00000  0.00000  8.00000  8.00000 30.22222                0
I(x1 * x2 * x3)       0.00000  0.00000  0.00000  0.00000  0.00000  0.00000  0.00000                8

round(solve (t(X)%*%X),3)
              (Intercept)    x1    x2    x3 I(x1^2) I(x2^2) I(x3^2) I(x1 * x2) I(x2 * x3) I(x1 * x3)
(Intercept)         0.124 0.000 0.000 0.000  -0.039  -0.039  -0.039      0.000      0.000      0.000
```

```
x1             0.000 0.068 0.000 0.000  0.000  0.000  0.000   0.000   0.000   0.000
x2             0.000 0.000 0.068 0.000  0.000  0.000  0.000   0.000   0.000   0.000
x3             0.000 0.000 0.000 0.068  0.000  0.000  0.000   0.000   0.000   0.000
I(x1^2)       -0.039 0.000 0.000 0.000  0.050  0.005  0.005   0.000   0.000   0.000
I(x2^2)       -0.039 0.000 0.000 0.000  0.005  0.050  0.005   0.000   0.000   0.000
I(x3^2)       -0.039 0.000 0.000 0.000  0.005  0.005  0.050   0.000   0.000   0.000
I(x1 * x2)     0.000 0.000 0.000 0.000  0.000  0.000  0.000   0.125   0.000   0.000
I(x2 * x3)     0.000 0.000 0.000 0.000  0.000  0.000  0.000   0.000   0.125   0.000
I(x1 * x3)     0.000 0.000 0.000 0.000  0.000  0.000  0.000   0.000   0.000   0.125
```

Le nombre d’expériences à réaliser est maintenant de 23. 9 points sont au centre.

Il apparaît que :

- La constante est indépendante des effets principaux, très peu liée aux effets quadratiques (-0,03) ;
- Les effets principaux sont totalement indépendants ;
- Les effets quadratiques liés à la constante (comme indiqué plus haut, 0,04) ;
- Les interactions sont indépendantes.

Pour améliorer le plan, il faudrait faire plus d’essais.

Exercice 13. Projet anti-gaspi (exemple réel, par Carla Rapini)

Un groupe d’étudiants-ingénieurs agroalimentaire mène un projet innovant et anti-gaspi. Ils souhaitent valoriser les fanes de légumes jetées en fin de marché dans une préparation culinaire type pesto. Le pesto est une spécialité italienne à base de basilic, d’huile d’olive, de parmesan, d’ail, de pignon de pin et de sel. Afin d’apporter une touche d’originalité, les étudiants voudraient faire un pesto à base de fanes de légumes (carotte ou radis), d’huile (olive ou pépin de raisin), de fruit à coque (noisette ou amande), de graines (sarrasin ou lin) et de fromage (feta ou parmesan).

Mais les étudiants ont une échéance et peu de moyens. Ils doivent donc limiter leur nombre d’essais à 8. Pour cela, ils décident de faire un plan d’expérience.

Les 5 facteurs étudiés prennent chacun deux modalités :

- Type de fanes (carotte, radis)
- Type d’huile (pépin de raisin, olive)
- Type de fruit à coque (noisette, amande)
- Type de graine (sarrasin, lin)
- Type de fromage (feta, parmesan)

1. **Combien d'essais sont nécessaires pour construire un plan complet à 5 facteurs à deux modalités ?**

Pour étudier un plan complet à 5 facteurs à 2 modalités, il faut $2^6 = 32$ essais.

2. **Écrire la matrice des essais du plan de base complet en 16 essais (le plan 2^4).**

A	B	C	D	E
1	1	1	1	1
1	1	1	-1	-1
1	1	-1	-1	1
1	1	-1	1	-1
1	-1	1	1	1
1	-1	1	-1	-1
1	-1	-1	-1	1
1	-1	-1	1	-1
-1	1	1	1	1
-1	1	1	-1	-1
-1	1	-1	-1	1

-1	1	-1	1	-1
-1	-1	1	1	1
-1	-1	1	-1	-1
-1 -1	-1	-1		1
-1 -1	-1	1		-1

3. Écrire la matrice des effets associée au modèle contenant tous les effets principaux et toutes les interactions d'ordre 2. En déduire la liste des confusions de ce plan puis la résolution du plan.

I	A	B	C	D	E	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE	ABC	ABD	ABE	ACD	ACE	ADE	BCD	BCE	CDE
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	1	-1	-1	1
1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	-1	1
1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	1
1	1	-1	1	1	1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	1
1	1	-1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	-1	1	1	1	1
1	1	-1	-1	-1	1	-1	-1	-1	1	1	1	-1	1	-1	-1	1	1	-1	1	-1	-1	-1	1	1
1	1	-1	-1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1
1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1
1	-1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1	-1	1	-1	1	1	1	1	-1	-1	-1	1
1	-1	1	-1	-1	1	-1	1	1	-1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	1	-1	1
1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	1	-1	1	1
1	-1	-1	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	-1	1
1	-1	-1	1	-1	-1	1	-1	1	1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	1	1	1
1	-1	-1	-1	-1	1	1	1	1	-1	1	1	-1	1	-1	-1	-1	-1	1	-1	1	1	-1	1	1
1	-1	-1	-1	1	-1	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	1

Les colonnes confondues (c'est-à-dire identiques) sont les suivantes : DE = C, CE = D, CD = E et CDE = I. Une seule interaction est confondue avec la constance : CDE. Cette interaction est d'ordre 3 donc le plan est de résolution III (la résolution du plan correspond à l'ordre de l'interaction la plus petite confondue avec la constante I).

4. Quel est le plan d'expérience optimal en 8 essais ?

```
library(FrF2)
plan <- FrF2(nruns=8, nfactors=5)
summary(plan)
The design itself:
   A  B  C  D  E
1  1  1 -1  1 -1
2 -1 -1 -1  1  1
3  1 -1 -1 -1 -1
4 -1  1 -1 -1  1
5  1  1  1  1  1
6 -1 -1  1  1 -1
7 -1  1  1 -1 -1
8  1 -1  1 -1  1
class=design, type= FrF2
```

	Type de fanes	Type d'huile	Type de fruit à coque	Type de graine	Type de fromage
1	carotte	olive	noisette	sarrasin	feta
2	radis	raisin	amande	sarrasin	feta
3	carotte	raisin	noisette	lin	feta
4	radis	olive	amande	lin	feta

5	radis	raisin	noisette	sarrasin	parmesan
6	carotte	olive	amande	lin	parmesan
7	radis	raisin	noisette	lin	parmesan
8	carotte	olive	amande	sarrasin	parmesan

PARTIE 2 : DEGUSTATION DES PRODUITS ANTI-GASPI

Une fois les essais réalisés, le groupe d'étudiants souhaite les faire déguster. Les 8 essais seront dégustés par 16 juges. Chaque juge dégustera chaque essai une fois et évaluera son goût et sa texture en mettant une note comprise entre 0 et 10.

- Goût (0 = fade ; 10=goûtu)
- Texture (0= granuleux ; 10= lisse)

L'ordre de dégustation varie d'un juge à l'autre pour éviter l'effet ordre et alors fausser les résultats.

5. Indiquer le principe de construction du plan.

Le produit (pesto), le juge et le rang de dégustation influencent la réponse. Les facteurs produit et rang prennent chacun 8 modalités et le facteur juge prend 16 modalités. On peut donc construire un plan en blocs incomplets équilibrés (BIE). Ainsi, tous les juges dégustent tous les produits un même nombre de fois dans un ordre différent.

6. Construire le plan.

La méthode pour construire le tableau est la suivante. Les juges, numérotés de 1 à 16, sont mis en ligne et le rang de dégustation des produits (de 1 à 8) en colonne. L'ordre de dégustation du juge 1 est le suivant : produit 1 – produit 2 – produit 8 – produit 2+1 – produit 8-1 – produit 2+2 – produit 8-2 – produit 2+3.

Ensuite, pour construire les ordres de dégustation des juges 2 à 8, on ajoute 1 à chaque produit par rapport à la ligne précédente. Par exemple, le juge 1 déguste le produit 1 au rang 1, le juge 2 déguste le produit 1+1 (= produit 2) au rang 1 et ainsi de suite. Quand on obtient le produit 8+1, on recommence au produit 1.

Pour le juge 9, on reprend l'ordre de dégustation du juge 1 et on inverse l'ordre. Puis pour les ordres de dégustation des juges 9 à 16, on ajoute 1 pour chaque produit par rapport à la ligne précédente comme vu auparavant. On obtient ainsi l'ordre de dégustation des 8 essais pour chaque juge.

		Rang							
		1	2	3	4	5	6	7	8
Juges	1	1	2	8	3	7	4	6	5
	2	2	3	1	4	8	5	7	6
	3	3	4	2	5	1	6	8	7
	4	4	5	3	6	2	7	1	8
	5	5	6	4	7	3	8	2	1
	6	6	7	5	8	4	1	3	2
	7	7	8	6	1	5	2	4	3
	8	8	1	7	2	6	3	5	4
	9	5	6	4	7	3	8	2	1
	10	6	7	5	8	4	1	3	2
	11	7	8	6	1	5	2	4	3
	12	8	1	7	2	6	3	5	4
	13	1	2	8	3	7	4	6	5
	14	2	3	1	4	8	5	7	6
	15	3	4	2	5	1	6	8	7
	16	4	5	3	6	2	7	1	8

7. Décrire les propriétés du plan.

Tous les juges dégustent tous les produits. Chaque produit est dégusté 2 fois au même rang. Chaque produit précède 2 fois et succède 2 fois le même produit. Le facteur produit est donc orthogonal au facteur juge et au facteur rang.

Exercice 14. Croissance de plantes en serre (exemple fictif, par P. Nuffer)

On veut réaliser une expérience en serre, pour regarder l'influence de différents facteurs sur la pousse de plantes.
On va s'intéresser à :

- L'apport d'eau, 10 ou 20 mL par jour
- La fertilisation, 2 ou 4 fois par semaine
- La température, 18 ou 20 degrés Celsius
- La photopériode, avec 14 ou 18 heures de lumière par jour
- Le pH du sol, avec un pH de 6,5 ou de 7,5
- La variété, entre deux variétés.

PARTIE 1. PLAN FRACTIONNAIRE

Dans un premier temps, on s'intéresse à une approche simplifiée. Pour cela, on regarde seulement les trois premiers facteurs, l'apport d'eau, de fertilisation et la température.

1. Ecrire la matrice des essais liée à cette situation.

Eau	Ferti	Temp
1	1	1
1	-1	-1
-1	1	-1
-1	-1	1

2. Ecrire la matrice des effets associée à ce modèle, contenant les effets principaux et les interactions. En déduire les confusions et la résolution du plan.

Matrice des effets

I	Eau	Ferti	Temp	Eau x Ferti	Eau x temp	Fert x Temp	Eau x Fert x Temp
1	1	1	1	1	1	1	1
1	1	-1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1
1	-1	-1	1	1	-1	-1	1

On peut observer que I=*Eau* x *Ferti* x *Temp* et Eau = *Ferti* x *Temp*, *Ferti* = *Eau* x *Temp* et Temp = *Eau* x *Ferti*
Seul *Eau* x *Ferti* x *Temp* est confondu avec I, le plan est donc de résolution III.

3. Quel modèle peut donc être étudié avec ce plan ? Quelles sont les conditions nécessaires ?

Le plan étant de résolution III, on peut seulement étudier le modèle avec les effets principaux. Pour le construire, on doit négliger les effets d'interaction d'ordre 2 et 3.

PARTIE 2. CONSTRUCTION D'UN PLAN OPTIMAL

Après cette première approche, on va prendre en compte les six facteurs pour construire un plan optimal.

4. Combien d'essais sont-ils nécessaires pour construire le plan complet ?

Il y a 6 facteurs à 2 modalités chacun. Ainsi, pour construire le plan complet il faut 2^6 = 64 expériences.

5. **Sur R, construire le plan complet et un plan optimal à 8 facteurs. (Utiliser les fonctions fac.design du package DoE.base, et optFederov du package AlgDesign).**

```
library(DoE.base)
plan <- fac.design(nlevels=2, factor.names = c("Eau", "Ferti", "Temp", "Lum", "pH", "Var")) #On
cree le plan complet en nomant les colonnes
plan

   Eau Ferti Temp Lum pH Var
1    1     1    2   2  1   1
2    2     2    2   2  2   2
3    2     1    2   1  1   1
4    1     2    1   1  1   2
5    2     1    2   1  2   1
6    1     2    2   1  1   2
7    1     2    2   2  2   2
8    2     1    2   1  1   2
9    1     2    2   1  1   1
10   2     2    1   2  1   2
11   2     1    1   1  2   2
12   1     1    2   1  2   2
13   1     2    2   1  2   2
14   1     2    1   2  1   2
15   2     1    2   1  2   2
16   2     2    1   2  1   1
17   1     1    1   1  2   2
18   2     2    1   2  2   2
19   1     1    2   1  1   1
20   2     2    2   1  2   2
21   2     2    1   2  2   1
22   1     2    2   2  1   1
23   2     1    2   2  2   2
24   2     1    1   2  2   1
25   1     2    1   1  1   1
26   2     1    1   2  2   2
27   1     2    2   2  2   1
28   1     1    1   2  1   2
29   1     2    1   2  1   1
30   1     1    1   2  2   1
31   1     1    2   2  2   1
32   2     1    1   2  1   2
33   1     1    2   2  1   2
34   2     1    1   1  1   2
35   2     2    2   2  1   2
36   1     1    2   1  1   2
37   1     2    1   2  2   2
38   2     2    1   1  1   2
39   2     2    2   2  1   1
40   2     2    1   1  2   2
41   1     2    1   1  2   1
42   1     1    1   2  2   2
43   1     1    1   2  1   1
44   1     1    2   2  2   2
45   1     2    2   2  1   2
46   2     2    2   1  1   1
47   2     1    2   2  2   1
48   1     2    2   1  2   1
49   1     2    1   1  2   2
50   1     1    1   1  1   1
51   2     2    2   1  1   2
52   1     1    2   1  2   1
53   1     2    1   2  2   1
54   2     2    1   1  2   1
55   1     1    1   1  1   2
56   2     2    2   2  2   1
57   1     1    1   1  2   1
58   2     2    1   1  1   1
59   2     1    2   2  1   1
60   2     1    1   2  1   1
61   2     1    2   2  1   2
62   2     1    1   1  1   1
63   2     2    2   1  2   1
64   2     1    1   1  2   1
class=design, type= full factorial

library(AlgDesign)
planOpt <- optFederov(~.,plan,nTrials=8,criterion="D")
planOpt

$D
```

```
[1] 1

$A
[1] 1

$Ge
[1] 1

$Dea
[1] 1

$design
   Eau Ferti Temp Lum pH Var
20   2     2    2   1  2   2
21   2     2    1   2  2   1
25   1     2    1   1  1   1
34   2     1    1   1  1   2
42   1     1    1   2  2   2
45   1     2    2   2  1   2
52   1     1    2   1  2   1
59   2     1    2   2  1   1

$rows
[1] 20 21 25 34 42 45 52 59
```

6. **Quel sont les différents degrés de libertés du modèle à 8 essais ?**

Il y a 7 degrés de libertés pour les paramètres, (1 pour la constante et 6 pour les facteurs). Il reste donc 1 degré de liberté pour la résiduelle.

7. **Calculer la matrice des effets associée au modèle puis la matrice de dispersion. Quelle propriété remarquez-vous ?**

```
options(contrasts = c("contr.sum", "contr.sum"))
Xopti <- model.matrix(~ . , planOpt$design)
t(Xopti)%*%Xopti

t(Xopti)%*%Xopti
            (Intercept) Eau1 Ferti1 Temp1 Lum1 pH1 Var1
(Intercept)           8    0      0     0    0   0    0
Eau1                  0    8      0     0    0   0    0
Ferti1                0    0      8     0    0   0    0
Temp1                 0    0      0     8    0   0    0
Lum1                  0    0      0     0    8   0    0
pH1                   0    0      0     0    0   8    0
Var1                  0    0      0     0    0   0    8
```

On peut voir que la matrice obtenue est orthogonale. De plus, on retrouve la matrice associée à un plan fractionnaire 2^{6-2}.

Exercice 15. Maximiser le poids d'un fromage (contexte inspiré d'un stage, données fictives, par Oriane David)

Une agricultrice fabrique des Saint Nectaire à sa ferme. Elle vend son fromage non affiné à la pièce mais la tarification se fait au poids du fromage. Elle souhaite donc avoir les Saint Nectaire les plus lourds possible pour vendre son Saint-Nectaire plus cher. Dans son laboratoire de fabrication elle peut jouer sur la taille des interstices de la cailleuse (de 5 à 20 mm), la pression de la mouleuse (entre 1,8 et 2,2 bar) , la durée du pressage (entre 8h et 12h) et la température de séchage (celle-ci-doit être entre 8 et 10 degrés). Elle décide d'allouer le moins de fromages possibles d'une fabrication à ces expériences (donc environ 30 fromages). Elle vous demande, en tant qu'ingénieur conseiller en production animale, de lui faire le plan de l'expérience.

1. **Quelles expériences lui conseillez-vous de réaliser ? Combien de fromages devrait-elle utiliser ? Commenter la qualité du plan obtenu.**

On lui conseille d'utiliser le plan suivant :

```
library(rsm)
planccd=ccd(4,coding=list(x1~(Temp-9)/1,  x2~(taille-12.5)/7.5,  x3~(duree-10)/2,  x4~(pression-2)/0.2))
planccd


planccd
   run.order std.order     Temp    taille    duree pression Block
1          1         8 10.00000 20.000000 12.00000 1.800000     1
```

```
2           2         14 10.00000  5.000000 12.00000 2.200000      1
3           3         17  9.00000 12.500000 10.00000 2.000000      1
4           4          1  8.00000  5.000000  8.00000 1.800000      1
5           5         13  8.00000  5.000000 12.00000 2.200000      1
6           6         15  8.00000 20.000000 12.00000 2.200000      1
7           7          3  8.00000 20.000000  8.00000 1.800000      1
8           8         20  9.00000 12.500000 10.00000 2.000000      1
9           9          9  8.00000  5.000000  8.00000 2.200000      1
10         10         10 10.00000  5.000000  8.00000 2.200000      1
11         11         18  9.00000 12.500000 10.00000 2.000000      1
12         12         19  9.00000 12.500000 10.00000 2.000000      1
13         13         11  8.00000 20.000000  8.00000 2.200000      1
14         14          4 10.00000 20.000000  8.00000 1.800000      1
15         15          2 10.00000  5.000000  8.00000 1.800000      1
16         16          6 10.00000  5.000000 12.00000 1.800000      1
17         17         12 10.00000 20.000000  8.00000 2.200000      1
18         18         16 10.00000 20.000000 12.00000 2.200000      1
19         19          5  8.00000  5.000000 12.00000 1.800000      1
20         20          7  8.00000 20.000000 12.00000 1.800000      1
21          1         12  9.00000 12.500000 10.00000 2.000000      2
22          2          6  9.00000 12.500000 14.38178 2.000000      2
23          3          4  9.00000 28.931677 10.00000 2.000000      2
24          4         10  9.00000 12.500000 10.00000 2.000000      2
25          5          3  9.00000 -3.931677 10.00000 2.000000      2
26          6          1  6.80911 12.500000 10.00000 2.000000      2
27          7         11  9.00000 12.500000 10.00000 2.000000      2
28          8          2 11.19089 12.500000 10.00000 2.000000      2
29          9          5  9.00000 12.500000  5.61822 2.000000      2
30         10          7  9.00000 12.500000 10.00000 1.561822      2
31         11          9  9.00000 12.500000 10.00000 2.000000      2
32         12          8  9.00000 12.500000 10.00000 2.438178      2

Data are stored in coded form using these coding formulas ...
x1 ~ (Temp - 9)/1
x2 ~ (taille - 12.5)/7.5
x3 ~ (duree - 10)/2
x4 ~ (pression - 2)/0.2
```

Elle devra alors utiliser 32 fromages soit une fabrication entière. Il y aura 8 essais de fromage qui seront les mêmes (essais moyens). On vérifie ensuite la qualité du plan.

```
X=model.matrix(~x1+x2+x3+x4+I(x1^2)+I(x2^2)+I(x3^2)+I(x4^2)+ TWI(x1,x2,x3,x4), data=planccd)
round(t(X)%*%X, 1)
round(solve(t(X)%*%X), 3)
```

```
 round(t(X)%*%X,1)
                                 (Intercept)   x1   x2   x3   x4 I(x1^2) I(x2^2) I(x3^2) I(x4^2)
 (Intercept)                            32.0  0.0  0.0  0.0  0.0    25.6    25.6    25.6    25.6
 x1                                      0.0 25.6  0.0  0.0  0.0     0.0     0.0     0.0     0.0
 x2                                      0.0  0.0 25.6  0.0  0.0     0.0     0.0     0.0     0.0
 x3                                      0.0  0.0  0.0 25.6  0.0     0.0     0.0     0.0     0.0
 x4                                      0.0  0.0  0.0  0.0 25.6     0.0     0.0     0.0     0.0
 I(x1^2)                                25.6  0.0  0.0  0.0  0.0    62.1    16.0    16.0    16.0
 I(x2^2)                                25.6  0.0  0.0  0.0  0.0    16.0    62.1    16.0    16.0
 I(x3^2)                                25.6  0.0  0.0  0.0  0.0    16.0    16.0    62.1    16.0
 I(x4^2)                                25.6  0.0  0.0  0.0  0.0    16.0    16.0    16.0    62.1
 TWI(x1, x2, x3, x4)x1:x2                0.0  0.0  0.0  0.0  0.0     0.0     0.0     0.0     0.0
 TWI(x1, x2, x3, x4)x1:x3                0.0  0.0  0.0  0.0  0.0     0.0     0.0     0.0     0.0
 TWI(x1, x2, x3, x4)x1:x4                0.0  0.0  0.0  0.0  0.0     0.0     0.0     0.0     0.0
 TWI(x1, x2, x3, x4)x2:x3                0.0  0.0  0.0  0.0  0.0     0.0     0.0     0.0     0.0
 TWI(x1, x2, x3, x4)x2:x4                0.0  0.0  0.0  0.0  0.0     0.0     0.0     0.0     0.0
 TWI(x1, x2, x3, x4)x3:x4                0.0  0.0  0.0  0.0  0.0     0.0     0.0     0.0     0.0
                                 TWI(x1, x2, x3, x4)x1:x2 TWI(x1, x2, x3, x4)x1:x3
 (Intercept)                                            0                        0
 x1                                                     0                        0
 x2                                                     0                        0
 x3                                                     0                        0
 x4                                                     0                        0
 I(x1^2)                                                0                        0
 I(x2^2)                                                0                        0
 I(x3^2)                                                0                        0
 I(x4^2)                                                0                        0
 TWI(x1, x2, x3, x4)x1:x2                              16                        0
 TWI(x1, x2, x3, x4)x1:x3                               0                       16
 TWI(x1, x2, x3, x4)x1:x4                               0                        0
 TWI(x1, x2, x3, x4)x2:x3                               0                        0
 TWI(x1, x2, x3, x4)x2:x4                               0                        0
 TWI(x1, x2, x3, x4)x3:x4                               0                        0
```

```
                          TWI(x1, x2, x3, x4)x1:x4 TWI(x1, x2, x3, x4)x2:x3
(Intercept)                                      0                        0
x1                                               0                        0
x2                                               0                        0
x3                                               0                        0
x4                                               0                        0
I(x1^2)                                          0                        0
I(x2^2)                                          0                        0
I(x3^2)                                          0                        0
I(x4^2)                                          0                        0
TWI(x1, x2, x3, x4)x1:x2                         0                        0
TWI(x1, x2, x3, x4)x1:x3                         0                        0
TWI(x1, x2, x3, x4)x1:x4                        16                        0
TWI(x1, x2, x3, x4)x2:x3                         0                       16
TWI(x1, x2, x3, x4)x2:x4                         0                        0
TWI(x1, x2, x3, x4)x3:x4                         0                        0
                          TWI(x1, x2, x3, x4)x2:x4 TWI(x1, x2, x3, x4)x3:x4
(Intercept)                                      0                        0
x1                                               0                        0
x2                                               0                        0
x3                                               0                        0
x4                                               0                        0
I(x1^2)                                          0                        0
I(x2^2)                                          0                        0
I(x3^2)                                          0                        0
I(x4^2)                                          0                        0
TWI(x1, x2, x3, x4)x1:x2                         0                        0
TWI(x1, x2, x3, x4)x1:x3                         0                        0
TWI(x1, x2, x3, x4)x1:x4                         0                        0
TWI(x1, x2, x3, x4)x2:x3                         0                        0
TWI(x1, x2, x3, x4)x2:x4                        16                        0
TWI(x1, x2, x3, x4)x3:x4                         0                       16
> round(solve(t(X)%*%X), 3)
                          (Intercept)    x1    x2    x3    x4 I(x1^2) I(x2^2) I(x3^2)
(Intercept)                     0.122 0.000 0.000 0.000 0.000  -0.028  -0.028  -0.028
x1                              0.000 0.039 0.000 0.000 0.000   0.000   0.000   0.000
x2                              0.000 0.000 0.039 0.000 0.000   0.000   0.000   0.000
x3                              0.000 0.000 0.000 0.039 0.000   0.000   0.000   0.000
x4                              0.000 0.000 0.000 0.000 0.039   0.000   0.000   0.000
I(x1^2)                        -0.028 0.000 0.000 0.000 0.000   0.025   0.003   0.003
I(x2^2)                        -0.028 0.000 0.000 0.000 0.000   0.003   0.025   0.003
I(x3^2)                        -0.028 0.000 0.000 0.000 0.000   0.003   0.003   0.025
I(x4^2)                        -0.028 0.000 0.000 0.000 0.000   0.003   0.003   0.003
TWI(x1, x2, x3, x4)x1:x2        0.000 0.000 0.000 0.000 0.000   0.000   0.000   0.000
TWI(x1, x2, x3, x4)x1:x3        0.000 0.000 0.000 0.000 0.000   0.000   0.000   0.000
TWI(x1, x2, x3, x4)x1:x4        0.000 0.000 0.000 0.000 0.000   0.000   0.000   0.000
TWI(x1, x2, x3, x4)x2:x3        0.000 0.000 0.000 0.000 0.000   0.000   0.000   0.000
TWI(x1, x2, x3, x4)x2:x4        0.000 0.000 0.000 0.000 0.000   0.000   0.000   0.000
TWI(x1, x2, x3, x4)x3:x4        0.000 0.000 0.000 0.000 0.000   0.000   0.000   0.000
                          I(x4^2) TWI(x1, x2, x3, x4)x1:x2 TWI(x1, x2, x3, x4)x1:x3
(Intercept)                -0.028                    0.000                    0.000
x1                          0.000                    0.000                    0.000
x2                          0.000                    0.000                    0.000
x3                          0.000                    0.000                    0.000
x4                          0.000                    0.000                    0.000
I(x1^2)                     0.003                    0.000                    0.000
I(x2^2)                     0.003                    0.000                    0.000
I(x3^2)                     0.003                    0.000                    0.000
I(x4^2)                     0.025                    0.000                    0.000
TWI(x1, x2, x3, x4)x1:x2    0.000                    0.062                    0.000
TWI(x1, x2, x3, x4)x1:x3    0.000                    0.000                    0.062
TWI(x1, x2, x3, x4)x1:x4    0.000                    0.000                    0.000
TWI(x1, x2, x3, x4)x2:x3    0.000                    0.000                    0.000
TWI(x1, x2, x3, x4)x2:x4    0.000                    0.000                    0.000
TWI(x1, x2, x3, x4)x3:x4    0.000                    0.000                    0.000
                          TWI(x1, x2, x3, x4)x1:x4 TWI(x1, x2, x3, x4)x2:x3
(Intercept)                                  0.000                    0.000
x1                                           0.000                    0.000
x2                                           0.000                    0.000
x3                                           0.000                    0.000
x4                                           0.000                    0.000
I(x1^2)                                      0.000                    0.000
I(x2^2)                                      0.000                    0.000
I(x3^2)                                      0.000                    0.000
I(x4^2)                                      0.000                    0.000
TWI(x1, x2, x3, x4)x1:x2                     0.000                    0.000
TWI(x1, x2, x3, x4)x1:x3                     0.000                    0.000
TWI(x1, x2, x3, x4)x1:x4                     0.062                    0.000
TWI(x1, x2, x3, x4)x2:x3                     0.000                    0.062
TWI(x1, x2, x3, x4)x2:x4                     0.000                    0.000
```

```
TWI(x1, x2, x3, x4)x3:x4                         0.000                         0.000
                         TWI(x1, x2, x3, x4)x2:x4 TWI(x1, x2, x3, x4)x3:x4
(Intercept)                                      0.000                         0.000
x1                                               0.000                         0.000
x2                                               0.000                         0.000
x3                                               0.000                         0.000
x4                                               0.000                         0.000
I(x1^2)                                          0.000                         0.000
I(x2^2)                                          0.000                         0.000
I(x3^2)                                          0.000                         0.000
I(x4^2)                                          0.000                         0.000
TWI(x1, x2, x3, x4)x1:x2                         0.000                         0.000
TWI(x1, x2, x3, x4)x1:x3                         0.000                         0.000
TWI(x1, x2, x3, x4)x1:x4                         0.000                         0.000
TWI(x1, x2, x3, x4)x2:x3                         0.000                         0.000
TWI(x1, x2, x3, x4)x2:x4                         0.062                         0.000
TWI(x1, x2, x3, x4)x3:x4                         0.000                         0.062
```

On remarque que les effets quadratiques sont légèrement confondus entre eux et avec la constante. Cette confusion est assez faible, mais cela peut avoir un effet sur l'analyse des résultats. Ce plan est globalement de bonne qualité.

2. **Finalement, ce plan lui paraît trop compliqué et trop long, car il faut régler les machines très précisément. L'agricultrice restreint donc les essais qu'elle voudrait faire : elle veut bien utiliser le même nombre de fromages et étudier certaines modalités des facteurs uniquement.**

   - **Interstice de 50, 100, 150 ou 200 mm uniquement**
   - **Pression de 1.8, 2 ou 2,2 bars uniquement**
   - **Durée de 8, 10 ou 12 heures**
   - **Température de 8,9 ou 10 degrés**

   De plus elle décide de supprimer les essais pour lesquels la pression de la mouleuse est de 1.8 bar et l'interstice est de 20 mm car les fromages ne rentrent pas dans les moules. Comparer la qualité du plan obtenu au précédent. Quel plan lui conseillez- vous alors ?

On créé d'abord un plan complet auquel on enlève les essais non réalisables.

```
library(DoE.base)
plan<- fac.design(nlevels=c(3,4,3,3),factor.names=list(x1=c(8,9,10),x2=c(5, 10, 15,
20),x3=c(8,10,12), x4=c(1.8, 2, 2.2)))
planmod=plan
for (i in 1:108){
  if ((planmod[i,2]==20) & (planmod[i,4]==1.8)){
    planmod=planmod[-i,]
    }
  }

library(AlgDesign)
planOpt <- optFederov(~.,planmod,98, nRepeats=50)
planOpt$design
```

```
planOpt$design
   x2 x1 x3  x4
2  15 10 12   2
4   5  8 10 1.8
14 20  8 10 2.2
16 15  9  8 1.8
18  5  9 12 2.2
20 15  8  8 2.2
21 10 10 12 1.8
29 10  8  8 2.2
30  5 10 10 2.2
33 15  8 12 2.2
40 10 10 10   2
43  5 10  8 2.2
51 10  8 10 1.8
59 20 10  8   2
60  5  8 12   2
62  5  8  8   2
65 10  9 10   2
67 10  8 12 1.8
```

```
70 15 10 10 2.2
71 20  9 10 2.2
72 15  9 10   2
76 20  9 12 2.2
77 15  9 10 1.8
78 15 10  8 1.8
81  5 10 12 1.8
83 10 10  8 2.2
84 10  9 12 2.2
85 20  8  8   2
89 15  8 12   2
91 20 10 12   2
93 10  9  8   2
99  5  9  8 1.8
```

L'algorithme produit plusieurs plans différents, le plan change à chaque fois que l'algorithme est lancé (même si on prend un nombre d'itérations plus grand).

On regarde ensuite la qualité du plan :

```
planOpt <- optFederov(~.,planmod,32, eval=T)
planOpt$design
X=model.matrix(~x1+x2+x3, data=planOpt$design)
round(t(X)%*%X,1)
round(solve(t(X)%*%X), 3)
(Intercept) x1.L x1.Q x2.L x2.Q x2.C x3.L x3.Q
(Intercept)        32.0 -1.4  0.8  1.8  0.0 -0.9 -1.4  0.8
x1.L               -1.4 11.0 -0.6  0.6  0.0 -0.3 -0.5  0.3
x1.Q                0.8 -0.6 10.3 -0.4  0.0  0.2  0.3 -0.2
x2.L                1.8  0.6 -0.4  8.0  0.4  0.0  0.6 -0.4
x2.Q                0.0  0.0  0.0  0.4  8.0  0.9  0.0  0.0
x2.C               -0.9 -0.3  0.2  0.0  0.9  8.0 -0.3  0.2
x3.L               -1.4 -0.5  0.3  0.6  0.0 -0.3 11.0 -0.6
x3.Q                0.8  0.3 -0.2 -0.4  0.0  0.2 -0.6 10.3
> round(solve(t(X)%*%X), 3)
            (Intercept)   x1.L   x1.Q   x2.L   x2.Q   x2.C   x3.L   x3.Q
(Intercept)       0.032  0.005 -0.003 -0.008  0.000  0.004  0.005 -0.003
x1.L              0.005  0.093  0.004 -0.009  0.000  0.004  0.005 -0.003
x1.Q             -0.003  0.004  0.098  0.005  0.000 -0.003 -0.003  0.002
x2.L             -0.008 -0.009  0.005  0.129 -0.007 -0.001 -0.009  0.005
x2.Q              0.000  0.000  0.000 -0.007  0.127 -0.014  0.000  0.000
x2.C              0.004  0.004 -0.003 -0.001 -0.014  0.128  0.004 -0.003
x3.L              0.005  0.005 -0.003 -0.009  0.000  0.004  0.093  0.004
x3.Q             -0.003 -0.003  0.002  0.005  0.000 -0.003  0.004  0.098
```

On voit que ce plan est beaucoup moins bon que le plan précédent pour le même nombre d'essais. En effet, il y a de légères confusions entre les effets principaux et on ne peut estimer les effets quadratiques car il s'agit d'un modèle avec variables catégorielles et non continues. Il sera donc très compliqué avec ce modèle d'arriver à de bonnes conclusions. Il serait alors nécessaire d'augmenter le nombre d'essais afin d'avoir de meilleurs résultats. En effet, le PPCM est 36. Il faudrait au minimum 36 essais afin d'avoir les facteurs principaux non confondus.

On calculera aussi la matrice relative au modèle avec interactions mais les résultats étant très long, on ne met pas ces résultats (Voir code). On observe la aussi des confusions entre les effets.

On voit alors que si l'on fait 36 essais et que l'on cherche à estimer les effets principaux seulement, alors ils ne sont pas confondus sauf pour x2 et x4 (voir code):

```
#Effets principaux seulement, 36 essais
planOpt <- optFederov(~.,planmod,36, nRepeats=50)
planOpt$design
X=model.matrix(~x1+x2+x3+x4, data=planOpt$design)
round(t(X)%*%X,1)
            (Intercept) x1.L x1.Q x2.L x2.Q x2.C x3.L x3.Q x4.L x4.Q
(Intercept)          36    0    0  0.0  0.0  0.0    0    0  0.0  0.0
x1.L                  0   12    0  0.0  0.0  0.0    0    0  0.0  0.0
x1.Q                  0    0   12  0.0  0.0  0.0    0    0  0.0  0.0
x2.L                  0    0    0  9.0  0.0  0.0    0    0  2.2 -1.6
x2.Q                  0    0    0  0.0  9.0  0.0    0    0  1.4  0.0
x2.C                  0    0    0  0.0  0.0  9.0    0    0 -0.3 -0.5
x3.L                  0    0    0  0.0  0.0  0.0   12    0  0.0  0.0
x3.Q                  0    0    0  0.0  0.0  0.0    0   12  0.0  0.0
x4.L                  0    0    0  2.2  1.4 -0.3    0    0 12.0  0.0
```

```
x4.Q                       0     0     0 -1.6  0.0 -0.5     0     0  0.0 12.0
round(solve(t(X)%*%X), 3)
            (Intercept)  x1.L  x1.Q   x2.L   x2.Q  x2.C  x3.L  x3.Q   x4.L   x4.Q
(Intercept)       0.028 0.000 0.000  0.000  0.000 0.000 0.000 0.000  0.000  0.000
x1.L              0.000 0.083 0.000  0.000  0.000 0.000 0.000 0.000  0.000  0.000
x1.Q              0.000 0.000 0.083  0.000  0.000 0.000 0.000 0.000  0.000  0.000
x2.L              0.000 0.000 0.000  0.120  0.004 0.000 0.000 0.000 -0.022  0.016
x2.Q              0.000 0.000 0.000  0.004  0.113 0.000 0.000 0.000 -0.014  0.000
x2.C              0.000 0.000 0.000  0.000  0.000 0.112 0.000 0.000  0.003  0.005
x3.L              0.000 0.000 0.000  0.000  0.000 0.000 0.083 0.000  0.000  0.000
x3.Q              0.000 0.000 0.000  0.000  0.000 0.000 0.000 0.083  0.000  0.000
x4.L              0.000 0.000 0.000 -0.022 -0.014 0.003 0.000 0.000  0.089 -0.003
x4.Q              0.000 0.000 0.000  0.016  0.000 0.005 0.000 0.000 -0.003  0.086

planOpt <- optFederov(~.,planmod,32, nRepeats=50)
planOpt$design
X=model.matrix(~.+ I(x1:x2)+I(x1:x3)+I(x1:x4)+I(x2:x3)+I(x4:x3)+I(x2:x4),data=planOpt$design)
round(t(X)%*%X,1)
round(solve(t(X)%*%X), 3)
```

La confusion entre x2 et x4 est probablement dû aux essais qui ont été enlevés car il s'agit des facteurs de la taille des interstices et de la pression.

Je lui conseillerais le deuxième type de plan si elle ne peut faire le premier. La confusion sur 32 essais est faible mais si elle peut faire 36 fromages cela serait idéal. Il faudrait cependant s'assurer qu'il n'y a pas d'effet quadratiques des facteurs sur le poids des fromages.

Exercice 16. Etude de la texture croustillante des cookies GRANOLA de la marque LU (exemple fictif, par Océane Guitton)

Partie 1 : Plans fractionnaires

Marine est en stage dans le département de recherche et développement chez LU, avec l'équipe qui est chargée de l'amélioration d'un des produits phare de la marque : le Cookie GRANOLA. Une analyse sensorielle par catégorisation, qui vient d'être réalisée par l'entreprise auprès des consommateurs, a révélé que ces derniers ne décrivent globalement pas les cookies GRANOLA comme croustillants tandis qu'ils décrivent la plupart des produits concurrents comme croustillants. Le problème est que le caractère croustillant est très apprécié par les consommateurs de cookies de la grande distribution. La marque LU a donc pour enjeu de se repositionner sur le marché. La marque LU charge donc l'équipe de recherche et développement de revoir la recette des cookies pour palier à ce problème.

Le maitre de stage de Marine, qui souhaite la laisser réfléchir seule au problème pour qu'elle puisse mettre à l'épreuve ses connaissances en planification expérimentale, la charge de préparer un plan d'étude qui dans la mesure du possible est réalisable sur le dernier mois de stage qu'il lui reste.

Marine commence donc par faire un état des lieux des différents facteurs qui selon elle peuvent influer sur la texture du cookie. Elle retient les facteurs suivants qui présentent chacun deux modalités :

- Type de sucre : (blanc, de canne)
- Nature de la farine : (bio, non bio)
- Température de cuisson (faible, forte)
- Temps de cuisson (court, long)

Marine envisage que les cookies issus de chaque essai d'expérience fassent par la suite l'objet d'une nouvelle analyse sensorielle par catégorisation. La variable réponse sera donc caractérisée par le pourcentage de cookies décrits par le mot « croustillant ».

1. **Marine pense dans un premier temps à un plan complet d'expériences. Quel sont les avantages et les inconvénients d'un tel type de plan ? Combien d'essais cela implique-t-il dans le cadre de l'étude de Marine ?**

Un plan complet d'expériences implique de faire toutes les expériences possibles, c'est-à-dire de considérer toutes les combinaisons possibles des modalités de chaque variable. La limite de ces plans est qu'ils présentent la plupart du temps un grand nombre d'essais qui ne sont pas toujours réalisables dans les limites de temps et de moyens.

Dans le cadre de l'étude de Marine, il y a 4 variables explicatives qui présentent chacune 2 modalités, ce qui implique 2^4 = 16 essais.

2. Selon vous, qu'est-ce qu'un « bon plan » ? Donner la formule de la variance des estimateurs (V(^β))

Un bon plan d'expérience est un plan qui va permettre d'avoir une estimation la plus précise possible de chacun des effets de chaque variable.

$$Var(\hat{\beta}) = (X'X)^{-1}\sigma^2$$

Sigma carré représente la variabilité résiduelle et dépend des résultats des expériences. Le membre $(X'X)^{-1}$ dépend uniquement du choix des expériences et donc de notre plan d'expérience : on va donc chercher trouver les expériences tel que $(X'X)^{-1}$ soit minimal pour avoir un bon plan d'expérience.

3. Construire sur le logiciel R à l'aide de la fonction FrF2 le plan complet d'expériences du modèle qui prend en compte toutes les interactions d'ordre 2 et calculer $(X'X)^{-1}$. Commenter.

```
library (FrF2)
plancomplet<FrF2(nruns=16,nfactors=4,factor.names=list(sucre=c("blanc","canne"),farine=c("bio","
non bio"),temperature=c("faible","forte"),temps=c("court","long")))
summary(plancomplet)
X<- model.matrix(~., data=plancomplet)
```

```
Call:
FrF2(nruns = 16, nfactors = 4, factor.names = list(sucre = c("blanc",
    "canne"), farine = c("bio", "non bio"),
    temperature = c("faible", "forte"), temps = c("court",
        "long")))

Experimental design of type  full factorial
16  runs

Factor settings:
  sucre  farine temperature temps
1 blanc     bio      faible court
2 canne non bio       forte  long

The design itself:
   sucre  farine temperature temps
1  canne     bio      faible court
2  blanc     bio      faible court
3  blanc non bio      faible court
4  blanc non bio      faible  long
5  blanc non bio       forte court
6  blanc     bio       forte  long
7  canne non bio       forte court
8  blanc     bio       forte court
9  canne non bio       forte  long
10 blanc non bio       forte  long
11 canne     bio      faible  long
12 blanc     bio      faible  long
13 canne non bio      faible court
14 canne     bio       forte  long
15 canne     bio       forte court
16 canne non bio      faible  long
class=design, type= full factorial
> X<- model.matrix(~.+sucre:farine+sucre:temperature+sucre:temps+farine:temperature+
    farine:temps+temperature:temps, data=plancomplet)
> t(X)%*%X
                     (Intercept) sucre1 farine1 temperature1 temps1 sucre1:farine1
(Intercept)                   16      0       0            0      0              0
sucre1                         0     16       0            0      0              0
farine1                        0      0      16            0      0              0
temperature1                   0      0       0           16      0              0
temps1                         0      0       0            0     16              0
sucre1:farine1                 0      0       0            0      0             16
sucre1:temperature1            0      0       0            0      0              0
sucre1:temps1                  0      0       0            0      0              0
farine1:temperature1           0      0       0            0      0              0
farine1:temps1                 0      0       0            0      0              0
temperature1:temps1            0      0       0            0      0              0
                     sucre1:temperature1 sucre1:temps1 farine1:temperature1 farine1:temps1
(Intercept)                            0             0                    0              0
sucre1                                 0             0                    0              0
farine1                                0             0                    0              0
temperature1                           0             0                    0              0
temps1                                 0             0                    0              0
sucre1:farine1                         0             0                    0              0
```

```
sucre1:temperature1                16          0                    0             0
sucre1:temps1                       0         16                    0             0
farine1:temperature1                0          0                   16             0
farine1:temps1                      0          0                    0            16
temperature1:temps1                 0          0                    0             0
                    temperature1:temps1
(Intercept)                         0
sucre1                              0
farine1                             0
temperature1                        0
temps1                              0
sucre1:farine1                      0
sucre1:temperature1                 0
sucre1:temps1                       0
farine1:temperature1                0
farine1:temps1                      0
temperature1:temps1                16
> solve(t(X)%*%X)
                    (Intercept) sucre1 farine1 temperature1 temps1 sucre1:farine1
(Intercept)              0.0625 0.0000  0.0000       0.0000 0.0000         0.0000
sucre1                   0.0000 0.0625  0.0000       0.0000 0.0000         0.0000
farine1                  0.0000 0.0000  0.0625       0.0000 0.0000         0.0000
temperature1             0.0000 0.0000  0.0000       0.0625 0.0000         0.0000
temps1                   0.0000 0.0000  0.0000       0.0000 0.0625         0.0000
sucre1:farine1           0.0000 0.0000  0.0000       0.0000 0.0000         0.0625
sucre1:temperature1      0.0000 0.0000  0.0000       0.0000 0.0000         0.0000
sucre1:temps1            0.0000 0.0000  0.0000       0.0000 0.0000         0.0000
farine1:temperature1     0.0000 0.0000  0.0000       0.0000 0.0000         0.0000
farine1:temps1           0.0000 0.0000  0.0000       0.0000 0.0000         0.0000
temperature1:temps1      0.0000 0.0000  0.0000       0.0000 0.0000         0.0000
                    sucre1:temperature1 sucre1:temps1 farine1:temperature1 farine1:temps1
(Intercept)                      0.0000        0.0000               0.0000         0.0000
sucre1                           0.0000        0.0000               0.0000         0.0000
farine1                          0.0000        0.0000               0.0000         0.0000
temperature1                     0.0000        0.0000               0.0000         0.0000
temps1                           0.0000        0.0000               0.0000         0.0000
sucre1:farine1                   0.0000        0.0000               0.0000         0.0000
sucre1:temperature1              0.0625        0.0000               0.0000         0.0000
sucre1:temps1                    0.0000        0.0625               0.0000         0.0000
farine1:temperature1             0.0000        0.0000               0.0625         0.0000
farine1:temps1                   0.0000        0.0000               0.0000         0.0625
temperature1:temps1              0.0000        0.0000               0.0000         0.0000
                    temperature1:temps1
(Intercept)                      0.0000
sucre1                           0.0000
farine1                          0.0000
temperature1                     0.0000
temps1                           0.0000
sucre1:farine1                   0.0000
sucre1:temperature1              0.0000
sucre1:temps1                    0.0000
farine1:temperature1             0.0000
farine1:temps1                   0.0000
temperature1:temps1              0.0625
```

Avec ce modèle, Marine peut estimer tous les effets principaux et toutes les interactions d'ordre 2. La variance des estimateurs des effets de chaque facteur est très faible : 0,0625. Ce plan est idéal. Cependant, ce plan d'expérience nécessite un certain temps. Si Marine opte pour un plan complet d'expériences, elle ne pourra pas assister à la mise en œuvre de la totalité des essais car la fin de son stage est imminente.

4. **Marine envisage donc par la suite de se limiter à 8 essais pour son plan d'expérience. Quel est le plan fractionnaire à construire dans le cadre de la situation de Marine ? Quelle est la particularité d'un tel type de plan ?**

Le plan fractionnaire à construire est le plan 2^{4-1} qui permet d'évaluer 4 facteurs à 2 modalités en 8 essais. La particularité d'un tel type de plan est que tous les facteurs sont orthogonaux. Les trois premiers facteurs (quantité sucre, quantité farine et température de cuisson) constituent un plan complet (en effet si on étudie 3 facteurs à deux modalités, le plan complet (2^3) présente 8 essais : c'est le plan de base.

5. **Construire la matrice des effets du modèle saturé dans le cadre de la construction du plan fractionnaire proposé à la question 2. Combien y a-t-il de générateurs d'alias ? Quels sont-ils ? Déterminer les confusions résultantes. Donner la résolution du plan.**

La matrice des effets est obtenue à partie du plan complet 2^3. La voici ci-dessous. On note A la quantité de sucre, B la quantité de farine, C la température de cuisson et D le temps de cuisson.

I	A	B	C	AB	AC	BC	ABC
1	1	1	1	1	1	1	1
1	1	1	-1	1	-1	-1	-1
1	1	-1	1	-1	1	-1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	1	-1	-1	1	-1
1	-1	1	-1	-1	1	-1	1
1	-1	-1	1	1	-1	-1	1
1	-1	-1	-1	1	1	1	-1

On confond donc D (temps de cuisson) avec l'interaction d'ordre le plus élevé : ABC. D = ABC ⇔ DD = ABCD ⇔ I = ABCD

Pour un plan fractionnaire 2^{4-1} il y a un générateur d'alias qui est : I = ABCD

On déduit les confusions résultantes de ce générateur d'alias : A = BCD, B = ACD et C = ABD puis AB = CD, AC = BD et BC = AD.

Quand il y a confusion d'effets, il est impossible de savoir ce qui est dû à l'un ou à l'autre. Ainsi, dans notre cas, pour pouvoir estimer les effets principaux A, B, C et D, il va falloir considérer les interactions d'ordre 3 ABC, BCD, ABD et ACD comme négligeables.

Puisque notre générateur d'alias est d'ordre 4, le plan 2^{4-1} est de résolution VI. Les effets principaux sont confondus avec des interactions d'ordre 3 et les interactions d'ordre 2 sont confondues avec d'autres interactions d'ordre 2. Nous allons donc pouvoir estimer les effets principaux sans problème mais nous ne pourrons pas estimer toutes interactions d'ordre 2 car certaines sont confondues entre elles.

6. **Marine propose le plan d'expérience suivant en 8 essais. Commenter le plan d'expériences proposé par Marine. Ecrire le modèle associé au plan d'expérience proposé par Marine.**

Essai	Quantité de sucre	Quantité de farine	Température de cuisson	Temps de cuisson
1	1	1	1	1
2	1	1	2	2
3	1	2	1	2
4	1	2	2	1
5	2	1	1	2
6	2	1	2	1
7	2	2	1	1
8	2	2	2	2

Le plan proposé par Marine est bien un plan fractionnaire 2^{4-1}. Les facteurs quantité de sucre, quantité de farine et température de cuisson constituent un plan complet. La température de cuisson est confondue avec l'interaction d'ordre 3 entre les trois autres facteurs. Ce plan est de résolution VI.

Le plan d'expérience proposé par Marine ne s'intéresse qu'aux effet principaux, le modèle est donc le suivant :

$$\forall i, j, k, l \qquad Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijkl}$$

$$\text{avec } \mathcal{L}(\varepsilon_{ijkl}) = \mathcal{N}(0, \sigma) \text{ et } \operatorname{Cov}(\varepsilon_{ijkl}, \varepsilon_{i'j'k'l'}) = 0 \quad \forall (i', j', k', l') \neq (i, j, k, l).$$

La variable réponse Y est le pourcentage de cookies décrits comme croustillants lors de l'analyse sensorielle.

7. **Ecrire la matrice des essais associée au plan d'expérience proposé par Marine.**

Quantité de sucre	Quantité de farine	Température de cuisson	Temps de cuisson
1	1	1	1
1	1	-1	-1
1	-1	1	-1

1	-1	-1	1
-1	1	1	-1
-1	1	-1	1
-1	-1	1	1
-1	-1	-1	-1

8. **Marine propose son plan d'expérience à son maitre de stage qui lui suggère d'utiliser le logiciel R afin d'évaluer la qualité de son plan avant de le mettre en œuvre et d'éventuellement le comparer à d'autres plans d'expériences suggérés par le logiciel. A l'aide du logiciel R, construire d'une part un plan fractionnaire à partir de la fonction FrF2, d'autre part le plan d'expériences proposé par Marine. Dans les deux cas, calculer $(X'X)^{-1}$.**

Voici ci-dessous les résultats obtenus pour le plan construit à partir de la fonction FrF2.

```
library (FrF2)
planR <- FrF2(nruns = 8, nfactors=4, factor.names=list(sucre = c("blanc","canne"),farine=c("bio"
,"non bio"),temperature=c("faible","forte"),temps=c("court","long")))> summary(planR)


Call:
FrF2(nruns = 8, nfactors = 4, factor.names = list(sucre = c("blanc",
    "canne"), farine = c("bio", "non bio"),
    temperature = c("faible", "forte"), temps = c("court",
        "long")))

Experimental design of type  FrF2
8  runs

Factor settings (scale ends):
  sucre  farine temperature temps
1 blanc     bio      faible court
2 canne non bio       forte  long

Design generating information:
$legend
[1] A=sucre       B=farine      C=temperature D=temps

$generators
[1] D=ABC


Alias structure:
$fi2
[1] AB=CD AC=BD AD=BC


The design itself:
  sucre  farine temperature temps
1 blanc     bio      faible court
2 blanc non bio      faible  long
3 canne     bio      faible  long
4 blanc non bio       forte court
5 canne     bio       forte court
6 canne non bio       forte  long
7 blanc     bio       forte  long
8 canne non bio      faible court
class=design, type= FrF2

X<- model.matrix(~., data=planR)
t(X)%*%X
             (Intercept) sucre1 farine1 temperature1 temps1
(Intercept)            8      0       0            0      0
sucre1                 0      8       0            0      0
farine1                0      0       8            0      0
temperature1           0      0       0            8      0
temps1                 0      0       0            0      8

solve(t(X)%*%X)
             (Intercept) sucre1 farine1 temperature1 temps1
(Intercept)        0.125  0.000   0.000        0.000  0.000
sucre1             0.000  0.125   0.000        0.000  0.000
farine1            0.000  0.000   0.125        0.000  0.000
temperature1       0.000  0.000   0.000        0.125  0.000
temps1             0.000  0.000   0.000        0.000  0.125
```

Voici ci-dessous les résultats obtenus à partir du plan d'expérience de Marine :

```
sucre <- as.factor(c(1,1,1,1,-1,-1,-1,-1))
farine <- as.factor(c(1,1,-1,-1,1,1,-1,-1))
temperature <- as.factor(c(1,-1,1,-1,1,-1,1,-1))
temps <- as.factor(c(1,-1,-1,1,-1,1,1,-1))
marine <-cbind.data.frame(sucre,farine,temperature,temps)
X<- model.matrix(~., data=marine)
t(X)%*%X
solve(t(X)%*%X)
             (Intercept) sucre1 farine1 temperature1 temps1
(Intercept)            8      0       0            0      0
sucre1                 0      8       0            0      0
farine1                0      0       8            0      0
temperature1           0      0       0            8      0
temps1                 0      0       0            0      8

             (Intercept) sucre1 farine1 temperature1 temps1
(Intercept)        0.125  0.00     0.00         0.00   0.00
sucre1             0.00   0.125    0.00         0.00   0.00
farine1            0.00   0.00     0.125        0.00   0.00
temperature1       0.00   0.00     0.00         0.125  0.00
temps1             0.00   0.00     0.00         0.00   0.125
```

9. **Commenter la qualité de chacun des deux plans obtenus. Commenter l'allure d'une matrice $(X'X)^{-1}$ issue d'essais choisis au hasard. Pensez-vous que Marine ait construit son plan en choisissant les essais au hasard ?**

Comme vu plus haut, regarder et analyser $(X'X)^{-1}$ permet de conclure sur la qualité du plan d'expérience. Les variances des estimateurs des effets de chaque facteur sont lisibles sur la diagonale. Les covariances entre deux facteurs sont lisibles en dehors de la diagonale : si la covariance n'est pas nulle, cela signifie que les deux facteurs ne sont pas indépendants.

On note que dans les deux cas, les variances sont minimales (0,125 = 1/8) ce qui signifie que les estimations des paramètres sont assez précises pour le plan d'expérience construit par la fonction FrF2 ainsi pour le plan d'expérience de Marine. De plus, les covariances sont nulles, tous les facteurs sont donc indépendants deux à deux dans ces plans d'expérience. On peut noter que la variance des estimateurs est plus élevée que dans le cadre du plan complet ce qui est normal puisqu'autant de facteurs sont estimés avec moins d'essais.

Ceci confirme donc le fait que Marine ait bien construit un plan d'expériences qui est optimal pour tester 4 facteurs en seulement 8 essais.

Marine a donc construit un plan idéal, ce qui est sans doute dû à ses connaissances en planification expérimentale : elle ne semble pas avoir choisi les essais au hasard.

Lorsque les essais sont choisis au hasard, les variances des estimateurs des effets de chaque facteur présentes sur la diagonale de $(X'X)^{-1}$ sont supérieures à celles des plans fractionnaires : avec des essaie choisis au hasard, les variances auraient été supérieures à 0,125. Par ailleurs, les covariances (que l'on trouve hors des diagonales de $(X'X)^{-1}$ ne sont pas toutes nulles dans le cas d'un plan d'expérience issu d'essais choisis au hasard car certains facteurs ne sont donc pas indépendants.

Pour conclure, il est primordial de bien choisir les essais pour avoir un bon plan d'expérience.

10. **À la suite de son analyse, Marine statue sur un plan d'expérience. Ce plan est réalisé et à la suite des 8 analyses sensorielles par catégorisation réalisées sur les cookies à la suite de chacun des 8 essais, les pourcentages de cookies décrits par l'appellation « croustillants » sont obtenus. Quelle est la fonction du logiciel R qui va permettre à Marine et son équipe de d'interpréter les facteurs qui ont un effet significatif sur la perception du « croustillant » du cookie ?**

La fonction AovSum du logiciel R permet de déterminer quels sont les facteurs estimés qui ont un effet significatif (p-value inférieure à 5%) sur la variable réponse à savoir le pourcentage de cookies décrits comme croustillants.

PARTIE 2 : PLANS SYMETRIQUES

Pour compléter l'étude et confirmer les résultats, l'entreprise LU décide d'organiser une séance d'analyse sensorielle complémentaire ou le jury cette fois ci n'est pas un jury de consommateurs mais un jury entrainé.

Chaque juge va déguster une fois 12 cookies : 8 cookies issus de chacun des 8 essais du plan d'expérience précédent ainsi que 4 cookies de marques concurrentes. Par ailleurs, le jury est composé de 12 juges. Le jury va devoir évaluer le « croustillant » du cookie sur une échelle allant de 0 (très mou) à 10 (très croustillant). La présentation des produits est monadique séquentielle (un cookie est présenté au juge lorsqu'il a fini d'évaluer le cookie précédent). De ce fait il faut tenir compte de l'effet rang.

11. Quels sont les facteurs qui peuvent influencer la note ? Proposer un type de plan de dégustation.

Trois facteurs peuvent influencer la note de dégustation le produit, le juge et le rang de dégustation. Il y a 12 juges, 12 cookies à déguster et par conséquent 12 rangs de dégustation possibles. Ainsi chacun des trois facteurs a 12 modalités. On peut donc proposer un plan de dégustation en carrés latins qui est un plan complet d'expériences.

12. Construire le plan de dégustation et décrire ses propriétés.

On propose le plan de dégustation suivant.

	Rang 1	Rang 2	Rang 3	Rang 4	Rang 5	Rang 6	Rang 7	Rang 8	Rang 9	Rang 10	Rang 11	Rang 12
Juge 1	1	2	3	4	5	6	7	8	9	10	11	12
Juge 2	2	3	4	5	6	7	8	9	10	11	12	1
Juge 3	3	4	5	6	7	8	9	10	11	12	1	2
Juge 4	4	5	6	7	8	9	10	11	12	1	2	3
Juge 5	5	6	7	8	9	10	11	12	1	2	3	4
Juge 6	6	7	8	9	10	11	12	1	2	3	4	5
Juge 7	7	8	9	10	11	12	1	2	3	4	5	6
Juge 8	8	9	10	11	12	1	2	3	4	5	6	7
Juge 9	9	10	11	12	1	2	3	4	5	6	7	8
Juge 10	10	11	12	1	2	3	4	5	6	7	8	9
Juge 11	11	12	1	2	3	4	5	6	7	8	9	10
Juge 12	12	1	2	3	4	5	6	7	8	9	10	11

Avec ce plan, tous les juges évaluent bien tous les produits, chaque produit est évalué 12 fois et chaque produit est évalué une fois à chaque rang. Les effets principaux orthogonaux : orthogonalité produit-rang, juge-rang et produit-juge. En revanche, le facteur produit est confondu avec l'interaction juge/rang, le facteur juge est confondu avec l'interaction produit/rang et le facteur rang est confondu avec l'interaction produit/juge.

13. Donner le modèle qui permettra d'étudier le plan.

L'étude du plan se fait par une analyse de la variance à 3 facteurs.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$

Avec Y la variable réponse, µ l'effet moyen, α_i l'effet du produit i, β_j l'effet du juge j et γ_k l'effet du rang k et ε_{ijk} les résidus.

14. Quel est la limite d'un tel plan d'expérience ? Peut-on proposer un type de plan alternatif ?

La limite d'un tel type de plan est qu'il ne tient pas compte d'un éventuel quatrième facteur pourrait influer sur la note de dégustation : le facteur produit précédent. Un plan des carrés latins spéciaux nommé plan de Williams permettrait de prendre en compte également l'effet du produit précédent.

Exercice 17. Recette d'houmous (exemple fictif, par Maxine Michaux)

PARTIE 1 : DETERMINER LA RECETTE IDEALE

Une entreprise veut créer une recette de houmous avec comme objectif ph de 4,5. Ce pH est le pH réglementaire pour éviter le développement des micro-organismes.

Le pH dépend uniquement des éléments suivants : pour 1000 kg de houmous, on mixe des pois-chiches, de l'eau, du tahini, du citron. Le temps de cuisson et la température interviennent également. Voici les intervalles de quantités, temps et températures, fixés à partir des recherches bibliographiques, pour la recette.

- Pois chiches : entre 600 et 650 kg
- Eau : 150 et 200 kg
- Tahini : 100 et 175 kg
- Citron : 75 et 100 kg
- Temps : 5h à 6h
- Température : 80° à 95°

On cherche alors à déterminer un plan d'expériences pour évaluer les variations du pH.

1. De quoi est constitué un plan composite centré ?

Un plan factoriel : les facteurs prennent 2 niveaux

Un point expérimental : au centre

Points axiaux : situés sur les axes de chacun des facteurs

2. Proposer un tel plan. Comment choisissons-nous α ?

On a un plan composite centré à 6 facteurs.

- Plan factoriel : 2^{6-1}. α = 2,378, n_α = 12, 15 points centre si orthogonalité, 9 si précision au centre. 59 points au total si orthogonalité. Nombre total d'expériences = 59

α est choisi pour respecter le critère d'orthogonalité.

3. Quelle ligne de code permettrai d'obtenir un tel plan ?

```
plan<-ccd(basis=~A+B+C+D+E,generators=F~A*B*C*D*E,randomize=FALSE, alpha="rotatable", coding=list(A~(Poischiche-625)/25, B~(Eau-175)/25, C~(tahini-137)/37,D~(Citron-87)/12,E~(temps-5)/1, F~(temperature-90)/10))
plan
   run.order std.order Poischiche      Eau    tahini   Citron    temps temperature Block
1          1         1   600.0000 150.0000 100.00000 75.00000 4.000000    80.00000     1
2          2         2   650.0000 150.0000 100.00000 75.00000 4.000000   100.00000     1
3          3         3   600.0000 200.0000 100.00000 75.00000 4.000000   100.00000     1
4          4         4   650.0000 200.0000 100.00000 75.00000 4.000000    80.00000     1
5          5         5   600.0000 150.0000 174.00000 75.00000 4.000000   100.00000     1
6          6         6   650.0000 150.0000 174.00000 75.00000 4.000000    80.00000     1
7          7         7   600.0000 200.0000 174.00000 75.00000 4.000000    80.00000     1
8          8         8   650.0000 200.0000 174.00000 75.00000 4.000000   100.00000     1
9          9         9   600.0000 150.0000 100.00000 99.00000 4.000000   100.00000     1
10        10        10   650.0000 150.0000 100.00000 99.00000 4.000000    80.00000     1
11        11        11   600.0000 200.0000 100.00000 99.00000 4.000000    80.00000     1
12        12        12   650.0000 200.0000 100.00000 99.00000 4.000000   100.00000     1
13        13        13   600.0000 150.0000 174.00000 99.00000 4.000000    80.00000     1
14        14        14   650.0000 150.0000 174.00000 99.00000 4.000000   100.00000     1
15        15        15   600.0000 200.0000 174.00000 99.00000 4.000000   100.00000     1
16        16        16   650.0000 200.0000 174.00000 99.00000 4.000000    80.00000     1
17        17        17   600.0000 150.0000 100.00000 75.00000 6.000000   100.00000     1
18        18        18   650.0000 150.0000 100.00000 75.00000 6.000000    80.00000     1
19        19        19   600.0000 200.0000 100.00000 75.00000 6.000000    80.00000     1
20        20        20   650.0000 200.0000 100.00000 75.00000 6.000000   100.00000     1
21        21        21   600.0000 150.0000 174.00000 75.00000 6.000000    80.00000     1
22        22        22   650.0000 150.0000 174.00000 75.00000 6.000000   100.00000     1
23        23        23   600.0000 200.0000 174.00000 75.00000 6.000000   100.00000     1
24        24        24   650.0000 200.0000 174.00000 75.00000 6.000000    80.00000     1
25        25        25   600.0000 150.0000 100.00000 99.00000 6.000000    80.00000     1
26        26        26   650.0000 150.0000 100.00000 99.00000 6.000000   100.00000     1
27        27        27   600.0000 200.0000 100.00000 99.00000 6.000000   100.00000     1
28        28        28   650.0000 200.0000 100.00000 99.00000 6.000000    80.00000     1
29        29        29   600.0000 150.0000 174.00000 99.00000 6.000000   100.00000     1
30        30        30   650.0000 150.0000 174.00000 99.00000 6.000000    80.00000     1
31        31        31   600.0000 200.0000 174.00000 99.00000 6.000000    80.00000     1
32        32        32   650.0000 200.0000 174.00000 99.00000 6.000000   100.00000     1
33        33        33   625.0000 175.0000 137.00000 87.00000 5.000000    90.00000     1
34        34        34   625.0000 175.0000 137.00000 87.00000 5.000000    90.00000     1
35        35        35   625.0000 175.0000 137.00000 87.00000 5.000000    90.00000     1
36        36        36   625.0000 175.0000 137.00000 87.00000 5.000000    90.00000     1
37         1         1   565.5396 175.0000 137.00000 87.00000 5.000000    90.00000     2
38         2         2   684.4604 175.0000 137.00000 87.00000 5.000000    90.00000     2
39         3         3   625.0000 115.5396 137.00000 87.00000 5.000000    90.00000     2
40         4         4   625.0000 234.4604 137.00000 87.00000 5.000000    90.00000     2
41         5         5   625.0000 175.0000  48.99867 87.00000 5.000000    90.00000     2
```

```
42          6          6   625.0000 175.0000 225.00133  87.00000 5.000000    90.00000      2
43          7          7   625.0000 175.0000 137.00000  58.45903 5.000000    90.00000      2
44          8          8   625.0000 175.0000 137.00000 115.54097 5.000000    90.00000      2
45          9          9   625.0000 175.0000 137.00000  87.00000 2.621586    90.00000      2
46         10         10   625.0000 175.0000 137.00000  87.00000 7.378414    90.00000      2
47         11         11   625.0000 175.0000 137.00000  87.00000 5.000000    66.21586      2
48         12         12   625.0000 175.0000 137.00000  87.00000 5.000000   113.78414      2
49         13         13   625.0000 175.0000 137.00000  87.00000 5.000000    90.00000      2
50         14         14   625.0000 175.0000 137.00000  87.00000 5.000000    90.00000      2
51         15         15   625.0000 175.0000 137.00000  87.00000 5.000000    90.00000      2
52         16         16   625.0000 175.0000 137.00000  87.00000 5.000000    90.00000      2

Data are stored in coded form using these coding formulas ...
A ~ (Poischiche - 625)/25
B ~ (Eau - 175)/25
C ~ (tahini - 137)/37
D ~ (Citron - 87)/12
E ~ (temps - 5)/1
F ~ (temperature - 90)/10
```

4. Combien d'essais réalisons-nous ? Combien estiment la précision du modèle ?

52 essais réalisés, 8 estiment la précision du modèle (les 8 combinaisons : 625, 175, 137, 87, 5, 90).

5. Les différentes recettes sont testées et le pH est observé. On obtient alors un vecteur Y. Y<-c (4.5, 4.6, 3.9, 4.7, 4.2, 3.6, 4.9, 4.7, 4.6, 4.5, 4.9, 5.2, 4.6, 4.3, 4.5, 4.8, 4.3, 4.3, 4.5, 4.2, 4.5,4.2,4.6,4.7,4.8,4.3,6,5.4,4.5,3.2,3.5,4.5,4.6,4.6,4.7,4.9,4.5,4.4,4.3,4.3,4.5,5.1,5.2,4.9,4.8,4.6,4.3,4.1,4.5,4.5,4.3 ,4.6)
Ce vecteur est concaténé au plan obtenu précédemment avec la fonction : plan2<-cbind(plan,Y)
On entre ensuite la formule suivante : reg <- rsm (Y~SO (A, B, C, D, E, F), data=plan2)
Que fait cette formule ? Commentez les résultats obtenus.

Elle montre quels effets sont significatifs sur la valeur de pH.

```
Call:
rsm(formula = Y ~ SO(A, B, C, D, E, F), data = plan2)

              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.5925212  0.1520478 30.2045  < 2e-16 ***
A           -0.0424309  0.0654651 -0.6481  0.52304    
B            0.1385243  0.0654651  2.1160  0.04491 *  
C           -0.0917250  0.0654651 -1.4011  0.17397    
D            0.0574062  0.0654651  0.8769  0.38924    
E           -0.0340696  0.0654651 -0.5204  0.60754    
F            0.0167226  0.0654651  0.2554  0.80056    
A:B          0.1375000  0.0761637  1.8053  0.08359 .  
A:C         -0.0312500  0.0761637 -0.4103  0.68523    
A:D         -0.0250000  0.0761637 -0.3282  0.74558    
A:E         -0.0687500  0.0761637 -0.9027  0.37568    
A:F          0.0125000  0.0761637  0.1641  0.87101    
B:C          0.0062500  0.0761637  0.0821  0.93528    
B:D          0.0625000  0.0761637  0.8206  0.41995    
B:E          0.0187500  0.0761637  0.2462  0.80764    
B:F         -0.0250000  0.0761637 -0.3282  0.74558    
C:D         -0.1937500  0.0761637 -2.5439  0.01782 *  
C:E         -0.0875000  0.0761637 -1.1488  0.26194    
C:F          0.0687500  0.0761637  0.9027  0.37568    
D:E         -0.0437500  0.0761637 -0.5744  0.57103    
D:F          0.1000000  0.0761637  1.3130  0.20161    
E:F          0.0687500  0.0761637  0.9027  0.37568    
A^2         -0.0355414  0.0565659 -0.6283  0.53573    
B^2         -0.0620579  0.0565659 -1.0971  0.28349    
C^2          0.0263304  0.0565659  0.4655  0.64578    
D^2          0.0705246  0.0565659  1.2468  0.22451    
E^2          0.0086527  0.0565659  0.1530  0.87970    
F^2         -0.0797356  0.0565659 -1.4096  0.17148    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.5595,     Adjusted R-squared:  0.064 
F-statistic: 1.129 on 27 and 24 DF,  p-value: 0.384

Analysis of Variance Table

Response: Y
                      Df Sum Sq  Mean Sq F value   Pr(>F)
FO(A, B, C, D, E, F)   6 1.4787 0.246446  1.3276 0.283450
TWI(A, B, C, D, E, F) 15 3.1000 0.206667  1.1133 0.395662
```

```
PQ(A, B, C, D, E, F)   6 1.0807 0.180109  0.9703 0.466165
Residuals             24 4.4551 0.185629
Lack of fit           17 4.2463 0.249785  8.3760 0.004129
Pure error             7 0.2088 0.029821
```

Seul l'ajout d'eau est significatif sur le pH. Il n'y a pas d'effet quadratique. L'interaction entre le citron et le tahini est significative.

6. **Commentez le tableau d'analyse de la variance.**

La p-value de l'erreur d'ajustement est de 0,004. On rejette l'hypothèse que le modèle s'ajuste bien.

PARTIE 2 : CONSTRUIRE UN PLAN OPTIMAL

On souhaite évaluer l'appréciation (de 1 à 10) de 4 houmous selon 3 facteurs :

- La recette : houmous avec pois chiches bio, houmous avec pois chiches bas de gamme, houmous avec pois chiches déjà précuits.
- L'emballage : opaque, transparent, poche en plastique.
- La marque : Fontaine Santé, Sabra, marque distributeur.

7. **Combien de facteurs et de modalités sont étudiées ?**

   On a ici 3 facteurs à différentes modalités : 4, 3 et 3.

8. **Combien d'essais sont nécessaires pour faire ce plan ?**

   $4X3^2$ = 36 essais

9. **Construire le plan complet sur R.**

```
library (DoE.base)
set.seed(123)  # pour toujours avoir les mêmes résultats
Design.1 <- fac.design(nlevels=c(4,3,3), factor.names= c("A","B","C"))
VARCOV <- solve(t(X)%*%X)
determinant <- det(VARCOV)
determinant
[1] 1.837575e-09
Design.1
   A B C
1  3 2 3
2  3 1 2
3  2 1 2
4  3 1 1
5  2 3 1
6  2 2 2
7  2 3 2
8  3 3 1
9  1 2 1
10 4 2 2
11 2 3 3
12 2 2 3
13 4 1 3
14 3 2 2
15 1 3 1
16 1 3 3
17 4 2 1
18 3 2 1
19 4 2 3
20 1 1 3
21 2 1 3
22 1 3 2
23 4 1 1
24 1 1 1
25 1 2 3
26 3 3 2
27 4 3 2
28 4 3 1
29 1 1 2
30 3 1 3
31 2 1 1
32 3 3 3
33 4 1 2
34 1 2 2
35 4 3 3
```

```
36 2 2 1
class=design, type= full factorial
```

Plan complet

10. **Construire le plan optimal en 10 essais. Ce plan est-il optimal ?**

```
set.seed(123)
Librairy(AlgDesign)
plan.1.Dopt<-optFederov(~.,Design.1,nTrials=10,criterion="D")
plan.1.Dopt
Xopti=model.matrix(~ . , plan.1.Dopt$design)
det(t(Xopti)%*%Xopti)
$D
[1] 0.3176595

$A
[1] 3.808594

$Ge
[1] 0.441

$Dea
[1] 0.282

$design
   A B C
3  2 1 2
5  2 3 1
10 4 2 2
12 2 2 3
18 3 2 1
22 1 3 2
24 1 1 1
26 3 3 2
30 3 1 3
35 4 3 3

$rows
 [1]  3  5 10 12 18 22 24 26 30 35
```

Plan optimal en 10 essais. Le déterminant est de 10368.

```
VARCOVopti <- solve(t(Xopti)%*%Xopti)
round(VARCOVopti, 2)
            (Intercept)   A.L   A.Q   A.C   B.L   B.Q   C.L  C.Q
(Intercept)        0.11  0.00  0.05  0.00 -0.02 -0.01  0.00 0.03
A.L                0.00  0.71  0.00  0.12 -0.12  0.21 -0.24 0.00
A.Q                0.05  0.00  0.44  0.00 -0.04 -0.03  0.00 0.05
A.C                0.00  0.12  0.00  0.38 -0.04  0.07 -0.08 0.00
B.L               -0.02 -0.12 -0.04 -0.04  0.32 -0.05  0.04 0.04
B.Q               -0.01  0.21 -0.03  0.07 -0.05  0.39 -0.07 0.02
C.L                0.00 -0.24  0.00 -0.08  0.04 -0.07  0.42 0.00
C.Q                0.03  0.00  0.05  0.00  0.04  0.02  0.00 0.29
```

La matrice n'est pas orthogonale, le plan n'est pas optimal.

11. **Trouver le plan optimal.**

PPCM(12,9)=36

```
set.seed(123)
plan.1.Dopt<-optFederov(~.,Design.1,nTrials=36,criterion="D")
plan.1.Dopt
Xopti=model.matrix(~ . , plan.1.Dopt$design)
det(t(Xopti)%*%Xopti)
VARCOVopti<-(solve(t(Xopti)%*%Xopti))
round(VARCOVopti,2)
            (Intercept)  A.L  A.Q  A.C  B.L  B.Q  C.L  C.Q
(Intercept)        0.03 0.00 0.00 0.00 0.00 0.00 0.00 0.00
A.L                0.00 0.11 0.00 0.00 0.00 0.00 0.00 0.00
A.Q                0.00 0.00 0.11 0.00 0.00 0.00 0.00 0.00
A.C                0.00 0.00 0.00 0.11 0.00 0.00 0.00 0.00
```

```
B.L                 0.00 0.00 0.00 0.00 0.08 0.00 0.00 0.00
B.Q                 0.00 0.00 0.00 0.00 0.00 0.08 0.00 0.00
C.L                 0.00 0.00 0.00 0.00 0.00 0.00 0.08 0.00
C.Q                 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.08
```

La matrice est orthogonale, le plan est optimal.

Exercice 18. Optimisation de la tendreté d'un pain de mie (exemple fictif, par Jean-Baptiste PETIT)

Cet énoncé est inspiré d'une expérience de stage mais les conditions d'expérience sont fictives.

Une entreprise de boulangerie industrielle souhaite mener une étude afin d'améliorer la consistance de son pain de mie. Leur objectif est d'améliorer la richesse en eau de la pâte après le premier passage dans le fermenteur afin que le produit final (après cuisson) soit plus tendre. La quantité d'eau absorbée par la pâte dépend de plusieurs facteurs :

- Le temps passé dans le fermenteur (10 à 15 minutes)
- Le taux d'humidité (de 80 à 90%)
- La chaleur du fermenteur (de 35 à 45°C)

La réalisation d'une expérience est un processus assez long, il est donc intéressant de mettre en place un Plan Composite Centré afin de réduire le nombre d'essais nécessaires.

1. **Donnez le modèle à utiliser afin d'étudier l'ensemble des facteurs**

On étudie le modèle complet avec les effets linéaires, quadratiques et les interactions.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_{11}X_1^2 + \beta_{22}X_2^2 + \beta_{33}X_3^2 + \beta_{12}X_1X_2 + \beta_{13}X_1X_3 + \beta_{23}X_2X_3 + \varepsilon$$

2. **Proposez un plan d'expériences réalisable en utilisant le logiciel R**

```
library(rsm)
plan <- ccd(3, coding=list (x1~(Temps-12.5)/2.5, x2~(Humidite-85)/5, x3~(Chaleur-40)/5))
plan
   run.order std.order     Temps Humidite  Chaleur Block
1          1         1 10.000000 80.00000 35.00000     1
2          2         2 15.000000 80.00000 35.00000     1
3          3         3 10.000000 90.00000 35.00000     1
4          4         4 15.000000 90.00000 35.00000     1
5          5         5 10.000000 80.00000 45.00000     1
6          6         6 15.000000 80.00000 45.00000     1
7          7         7 10.000000 90.00000 45.00000     1
8          8         8 15.000000 90.00000 45.00000     1
9          9         9 12.500000 85.00000 40.00000     1
10        10        10 12.500000 85.00000 40.00000     1
11        11        11 12.500000 85.00000 40.00000     1
12        12        12 12.500000 85.00000 40.00000     1
13         1         1  7.935645 85.00000 40.00000     2
14         2         2 17.064355 85.00000 40.00000     2
15         3         3 12.500000 75.87129 40.00000     2
16         4         4 12.500000 94.12871 40.00000     2
17         5         5 12.500000 85.00000 30.87129     2
18         6         6 12.500000 85.00000 49.12871     2
19         7         7 12.500000 85.00000 40.00000     2
20         8         8 12.500000 85.00000 40.00000     2
21         9         9 12.500000 85.00000 40.00000     2
22        10        10 12.500000 85.00000 40.00000     2

Data are stored in coded form using these coding formulas ...
x1 ~ (Temps - 12.5)/2.5
x2 ~ (Humidite - 85)/5
x3 ~ (Chaleur - 40)/5
```

3. **Etudier la qualité de ce plan**

On calcule la matrice $(X'X)^{-1}$ sur R :

```
X <- model.matrix(~x1+x2+x3+I(x1^2)+I(x2^2)+I(x3^2)+I(x1*x2)+I(x1*x3)+I(x2*x3),data=plan)
round(solve(t(X)%*%X),3)
            (Intercept)    x1    x2    x3 I(x1^2) I(x2^2) I(x3^2) I(x1 * x2) I(x1 * x3) I(x2 * x3)
(Intercept)       0.124 0.000 0.000 0.000  -0.039  -0.039  -0.039      0.000      0.000      0.000
x1                0.000 0.068 0.000 0.000   0.000   0.000   0.000      0.000      0.000      0.000
x2                0.000 0.000 0.068 0.000   0.000   0.000   0.000      0.000      0.000      0.000
x3                0.000 0.000 0.000 0.068   0.000   0.000   0.000      0.000      0.000      0.000
I(x1^2)          -0.039 0.000 0.000 0.000   0.050   0.005   0.005      0.000      0.000      0.000
```

```
I(x2^2)        -0.039 0.000 0.000 0.000  0.005   0.050   0.005     0.000     0.000     0.000
I(x3^2)        -0.039 0.000 0.000 0.000  0.005   0.005   0.050     0.000     0.000     0.000
I(x1 * x2)      0.000 0.000 0.000 0.000  0.000   0.000   0.000     0.125     0.000     0.000
I(x1 * x3)      0.000 0.000 0.000 0.000  0.000   0.000   0.000     0.000     0.125     0.000
I(x2 * x3)      0.000 0.000 0.000 0.000  0.000   0.000   0.000     0.000     0.000     0.125
```

Les effets principaux sont indépendants des autres effets, tout comme l'interaction. Seuls les effets quadratiques sont légèrement liés à la constante, mais cette liaison est faible. Le plan est donc assez satisfaisant.

4. **Dans la réalité, un essai est long à mettre en place. On souhaite donc ne réaliser que 10 essais. Quel nouveau plan peut être mis en place ? Ce plan est-il intéressant ?**

Il faut mettre en place un plan optimal en utilisant l'algorithme de Federov pour un plan en 10 essais.

```
library(AlgDesign)
desL <- optFederov(~x1+x2+x3, data=plan, nTrials=10, eval=TRUE)
desL
$design
   run.order std.order      Temps Humidite  Chaleur Block
2          2         2 15.000000 80.00000 35.00000     1
3          3         3 10.000000 90.00000 35.00000     1
5          5         5 10.000000 80.00000 45.00000     1
8          8         8 15.000000 90.00000 45.00000     1
13         1         1  7.935645 85.00000 40.00000     2
14         2         2 17.064355 85.00000 40.00000     2
15         3         3 12.500000 75.87129 40.00000     2
16         4         4 12.500000 94.12871 40.00000     2
17         5         5 12.500000 85.00000 30.87129     2
18         6         6 12.500000 85.00000 49.12871     2
```

On étudie la qualité du plan par rapport au plan précédent :

```
Xopti=model.matrix(~x1+x2+x3+I(x1^2)+I(x2^2)+I(x3^2)+I(x1*x2)+I(x1*x3)+I(x2*x3), desL$design)
round(solve(t(Xopti)%*%Xopti),3)
            (Intercept)    x1    x2    x3 I(x1^2) I(x2^2) I(x3^2) I(x1 * x2) I(x1 * x3) I(x2 * x3)
(Intercept)        38.5  0.00  0.00  0.00 -12.000 -12.000 -12.000       0.00       0.00       0.00
x1                  0.0  0.15  0.00  0.00   0.000   0.000   0.000       0.00       0.00      -0.15
x2                  0.0  0.00  0.15  0.00   0.000   0.000   0.000       0.00      -0.15       0.00
x3                  0.0  0.00  0.00  0.15   0.000   0.000   0.000      -0.15       0.00       0.00
I(x1^2)           -12.0  0.00  0.00  0.00   3.780   3.735   3.735       0.00       0.00       0.00
I(x2^2)           -12.0  0.00  0.00  0.00   3.735   3.780   3.735       0.00       0.00       0.00
I(x3^2)           -12.0  0.00  0.00  0.00   3.735   3.735   3.780       0.00       0.00       0.00
I(x1 * x2)          0.0  0.00  0.00 -0.15   0.000   0.000   0.000       0.40       0.00       0.00
I(x1 * x3)          0.0  0.00 -0.15  0.00   0.000   0.000   0.000       0.00       0.40       0.00
I(x2 * x3)          0.0 -0.15  0.00  0.00   0.000   0.000   0.000       0.00       0.00       0.40
```

Les effets principaux ne sont pas orthogonaux avec les interactions, ce plan est donc de relativement mauvaise qualité, il est probable que les 22 essais du plan précédent doivent être faits.

Exercice 19. Plan pour obtenir un gout de bière optimale (exemple fictif, par N. Boutrand)

Votre frère a décidé de se lancer dans la confection de bières artisanales. Il veut à tout prix essayer de reproduire le goût de sa bière préférée, se rendant compte que sur le long terme, cela lui coûtera beaucoup moins cher que de l'acheter en supermarché. Il veut donc déterminer l'optimum des facteurs pour lesquels sa bière ressemblera le plus à sa bière préférée. Les facteurs pris en compte ici sont le malt entre 4 et 6kg pour 20L de bière, et la quantité de houblon entre 50 et 100g.

Une note sera donnée à chaque bière en fonction de la ressemblance, plus la note est élevée plus l'essai ressemble à l'originale. Ainsi, la variable d'intérêt Y ici est la note reçue pour chaque essai.

On recherche donc la combinaison {X1, X2} des facteurs conduisant à une valeur maximum de la réponse Y.

1. **Quel est le modèle du plan que nous souhaitons étudier ?**

Le modèle est le suivant :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2 + \beta_{12} X_1 X_2 + \varepsilon$$

Cette formule comporte chaque facteur, les effets quadratiques de ces facteurs et la confusion entre eux. Il contient également le résidus epsilon. Nous allons donc évaluer ces effets par la suite.

2. **Quels résultats obtient-on lorsque l'on crée le plan sur R Studio ? Analysez le modèle.**

```
library(rsm)
planbiere<-ccd(2, n0=2, randomize=FALSE, coding=list (x1~(Houblon-75)/25, x2~(Malt-5)/1))
planbiere
X <- model.matrix(~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data=planbiere)
solve(t(X)%*%X) # qualité du modèle
            (Intercept)    x1    x2  I(x1^2)  I(x2^2) I(x1 * x2)
(Intercept)       0.250 0.000 0.000 -0.12500 -0.12500       0.00
x1                0.000 0.125 0.000  0.00000  0.00000       0.00
x2                0.000 0.000 0.125  0.00000  0.00000       0.00
I(x1^2)          -0.125 0.000 0.000  0.15625  0.03125       0.00
I(x2^2)          -0.125 0.000 0.000  0.03125  0.15625       0.00
I(x1 * x2)        0.000 0.000 0.000  0.00000  0.00000       0.25
```

Nous pouvons dire que cette matrice est optimale, car la constante est indépendante des effets principaux, légèrement liée aux effets quadra. De plus, les effets principaux et les interactions sont indépendants des autres effets. Les coefficients sont estimés sans qu'il y ait de liaison avec les autres.

3. **On estime que les bières ont été notées sur 10, avec les notes suivantes : 2,6,7,4,8,3,9,4,4,1,7,6. Commentez les effets de ce plan ?**

```
set.seed(1234)
Y <- c(2,6,7,4,8,3,9,4,4,1,7,6)
rsm.biere <- rsm(Y~SO(x1,x2),data=planbiere)
summary(rsm.biere)

Call:
rsm(formula = Y ~ SO(x1, x2), data = planbiere)

            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  6.00000    1.09592  5.4749 0.001551 **
x1          -0.75888    0.77493 -0.9793 0.365254   
x2          -0.15533    0.77493 -0.2004 0.847755   
x1:x2       -1.75000    1.09592 -1.5968 0.161415   
x1^2         0.31250    0.86640  0.3607 0.730690   
x2^2        -1.68750    0.86640 -1.9477 0.099373 . 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Nous observons tout d'abord que le R carré est de 0.569, il n'est donc pas très bon.

Nous voyons que toutes les plus-values des effets et interactions sont supérieures à 5%, il n'y a donc aucun effet significatif des effets linéaires, quadratiques et des interactions sur le modèle.

De plus, l'erreur d'ajustement (*lack of fit*) est supérieure à 5%, elle n'est donc pas du même ordre de grandeur que l'erreur pure. Le model n'est donc pas optimal, il n'est pas bien choisi.

4. **Quels sont les valeurs optimales de houblons et de malt ? que peut-on déduire ?**

```
Multiple R-squared:  0.5692,	Adjusted R-squared:  0.2103 
F-statistic: 1.586 on 5 and 6 DF,  p-value: 0.2938

Analysis of Variance Table

Response: Y
            Df  Sum Sq Mean Sq F value Pr(>F)
FO(x1, x2)   2  4.8003  2.4001  0.4996 0.6300
TWI(x1, x2)  1 12.2500 12.2500  2.5499 0.1614
PQ(x1, x2)   2 21.0417 10.5208  2.1900 0.1931
Residuals    6 28.8247  4.8041               
Lack of fit  3 14.8247  4.9416  1.0589 0.4818
Pure error   3 14.0000  4.6667               

Stationary point of response surface:
        x1         x2 
 0.4426642 -0.2755533 

Stationary point in original units:
  Houblon      Malt 
86.066606  4.724447 

Eigenanalysis:
eigen() decomposition
$values
[1]  0.6412682 -2.0162682

$vectors
         [,1]      [,2]
x1 -0.9361027 0.3517267
x2  0.3517267 0.9361027
```

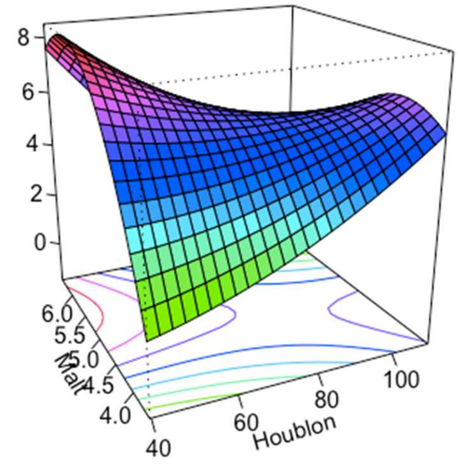
Même si le modèle n'est pas bien conçu, nous pouvons quand même analyser les résultats suivants.

Nous voyons que les optimums pour le houblon et le malt sont respectivement : 86.06 et 4.72. Ce sont donc les quantités qu'il faut mettre pour obtenir la note maximale.

Nous voyons également que les *eigenvalues* sont positives et négatives, on peut donc en déduire de ces résultats que l'optimum est un point selle,

Nous pouvons construire un graphique de répartition pour confirmer cette dernière affirmation :

```
contour(rsm.biere,~x1+x2,image=TRUE)
persp(rsm.biere,~x1+x2,col=rainbow(50), contours="colors")
```



Nous voyons ici qu'en effet l'optimum est bien un point selle.

PARTIE 2 : AJOUT DE FACTEURS (PROPOSE PAR N. BOUTRAND)

Au vu des résultats du premier modèle, et après quelques recherches, votre frère s'est finalement rendu compte que d'autres facteurs rentraient en compte dans le goût final de la bière. Cependant certains de ces facteurs possèdent plus de deux modalités.

Les quatre nouveaux facteurs à prendre en compte sont donc les suivants, suivis de leur modalité :

- *Taux d'atténuation de la levure (capacité de transformer du sucre en alcool en pourcentage) : 70%, 80%, 90%*
- *Temps de fermentation (en semaine) : 2, 3, 4, 5 semaines*
- *Température de fermentation (en degrés Celsius) : 19º, 21º*
- *Resucrage (sucre rajouté au moment de la mise en bouteille, en g/L) : 6, 8*

Les brassages durant en moyenne 1 mois, et étant un peu couteux, il est important de minimiser le nombre d'essais à 8.

5. **Combien d'essais faut-il réaliser pour que le plan soit complet ?**

Nous avons 4 facteurs à successivement 3,4,2 et 2 modalités. Ainsi nous pouvons écrire la formule suivante :

Essais minimum = $3*4*2^2$ = 48 essais

Avec la commande suivante, en créant le modèle sur R Studio nous avons obtenu ce même résultat :

```
library(DoE.base)
plan1 <-fac.design(nlevels=c(3,4,2,2),factor.names=(c("Atténuation", "TpsFerm", "TempFerm",
"Resucrage")))
```

```
Atténuation TpsFerm TempFerm Resucrage
1            2        1         2          2
2            3        2         1          2
3            3        3         2          2
4            2        4         2          1
5            1        2         2          2
6            3        2         2          1
7            1        3         2          1
8            3        1         2          2
9            1        2         1          2
10           1        1         1          1
11           1        3         1          1
12           3        2         2          2
```

```
13           1       2        2         1
14           2       4        1         2
15           1       3        1         2
16           3       4        1         1
17           1       4        2         2
18           2       2        1         2
19           3       3        2         1
20           1       1        2         2
21           2       2        1         1
22           2       2        2         1
23           2       3        1         1
24           3       3        1         2
25           3       2        1         1
26           3       4        2         1
27           3       1        2         1
28           1       4        1         1
29           1       2        1         1
30           2       3        1         2
31           1       1        2         1
32           2       4        1         1
33           3       1        1         1
34           2       1        1         1
35           2       1        2         1
36           2       3        2         2
37           3       3        1         1
38           2       4        2         2
39           3       4        1         2
40           2       2        2         2
41           1       3        2         2
42           1       4        1         2
43           3       1        1         2
44           1       4        2         1
45           3       4        2         2
46           2       3        2         1
47           2       1        1         2
48           1       1        1         2
class=design, type= full factorial
```

6. **Comment construire sur R Studio un plan optimal avec 8 essais avec le package AlgDesign ? Nous sommes dans un modèle d'analyse de variance à quatre facteurs, quels sont les degrés de liberté associés à chacun de ces facteurs ?**

Voici le résultat obtenu sur R Studio avec la fonction suivante :

```
set.seed(123)
library(AlgDesign)
planOpt<-optFederov(~.,data=plan1,nTrials=8)
planOpt
$D
[1] 0.4055974

$A
[1] 3.454861

$Ge
[1] 0.5

$Dea
[1] 0.368

$design
   Atténuation TpsFerm TempFerm Resucrage
4            2       4        2         1
15           1       1        1         1
20           1       3        1         2
23           3       3        2         1
26           1       2        2         2
28           3       4        1         2
31           2       2        1         1
33           2       1        2         2

$rows
[1]  4 15 20 23 26 28 31 33
```

Pour connaitre les degrés de liberté ddl, il faut d'abord calculer le nombre de paramètres à estimer :

1+1*(3-1) +2*(2-1) +1*(4-1) = 1+2+2+3 = 8 ddl

Comme nous faisons 8 essais, nous n'avons plus de ddl pour estimer la résiduelle (8essais-8ddl = 0). Ainsi il en reste 0, donc le modèle est saturé.

7. Comparez ce plan à un plan complet à l'aide des matrices de dispersion. Que pouvons-nous en conclure ?

Voici la matrice de dispersion de notre plan à 8 essais :

```
set.seed(1234)
planOpt<-optFederov(~Atténuation+TpsFerm+TempFerm+Resucrage,data=plan1,nTrials=8)
MatriceOpt<-model.matrix(~Atténuation+TpsFerm+TempFerm+Resucrage,planOpt$design)
Mat<-solve((t(MatriceOpt)%*%MatriceOpt))
round(Mat,2)
```

```
             (Intercept) Atténuation.L Atténuation.Q TpsFerm.L TpsFerm.Q TpsFerm.C TempFerm1 Resucrage1
(Intercept)         0.15         -0.12         -0.07     -0.07     -0.08      0.04      0.00       0.00
Atténuation.L      -0.12          0.67          0.00      0.42      0.24     -0.21      0.00       0.00
Atténuation.Q      -0.07          0.00          0.67      0.00      0.41      0.00      0.00       0.00
TpsFerm.L          -0.07          0.42          0.00      0.77      0.15     -0.13      0.00       0.00
TpsFerm.Q          -0.08          0.24          0.41      0.15      0.83     -0.07      0.00       0.00
TpsFerm.C           0.04         -0.21          0.00     -0.13     -0.07      0.57      0.00       0.00
TempFerm1           0.00          0.00          0.00      0.00      0.00      0.00      0.17      -0.08
Resucrage1          0.00          0.00          0.00      0.00      0.00      0.00     -0.08       0.17
```

Nous voyons qu'il y a des confusions pour tous les facteurs, à part 3. Nous pouvons en déduire que le plan n'est pas très précis et ne donnera pas de bons résultats.

Voici celle du plan complet à 48 essais :

```
set.seed(12345)
planC<-optFederov(~Atténuation+TpsFerm+TempFerm+Resucrage,data=plan1,nTrials=48)
MatriceC<-model.matrix(~Atténuation+TpsFerm+TempFerm+Resucrage,planC$design)
Mat2<-solve((t(MatriceC)%*%MatriceC))
round(Mat2,2)
```

```
             (Intercept) Atténuation.L Atténuation.Q TpsFerm.L TpsFerm.Q TpsFerm.C TempFerm1 Resucrage1
(Intercept)         0.02          0.00          0.00      0.00      0.00      0.00      0.00       0.00
Atténuation.L       0.00          0.06          0.00      0.00      0.00      0.00      0.00       0.00
Atténuation.Q       0.00          0.00          0.06      0.00      0.00      0.00      0.00       0.00
TpsFerm.L           0.00          0.00          0.00      0.08      0.00      0.00      0.00       0.00
TpsFerm.Q           0.00          0.00          0.00      0.00      0.08      0.00      0.00       0.00
TpsFerm.C           0.00          0.00          0.00      0.00      0.00      0.08      0.00       0.00
TempFerm1           0.00          0.00          0.00      0.00      0.00      0.00      0.02       0.00
Resucrage1          0.00          0.00          0.00      0.00      0.00      0.00      0.00       0.02
```

Nous voyons ici que quasiment aucuns des facteurs ne sont corrélés entre eux ou avec l'intercept (la constante). Ce plan est donc un bon plan.

Nous pouvons en conclure que le plan complet, avec 48 essais, est mieux adapté pour les essais de mon frère que celui à 8 essais. Il y a en effet dans ce plan trop peu d'essais et les résultats seraient faussés.

Exercice 20. Recette secrète du fondant Baulois (exemple réel, par Alix MARC)

Le fondant Baulois est une spécialité régionale et mais c'est avant tout un fondant au chocolat au goût et à la texture inimitable. Le secret est bien gardé par l'entreprise depuis 35 ans. Notre but est de réussir à trouver la recette s'approchant le plus de ce fondant unique donc on ne connait que la liste d'ingrédients figurant sur l'étiquette. Pour fixer les variables, nous nous sommes inspirés de nombreuses recettes se trouvant tout de même sur internet. On remarque que les quantités d'œuf, de chocolat et de farine ne varient pas donc on les fixe, pour un petit gâteau, à 1oeuf, 40g de chocolat et 5g de farine. Les quantités de sucre et du beurre sont en revanche variables. On a aussi noté deux autres facteurs liés aux ingrédients comme la proportion de chocolat noir/chocolat au lait, la proportion de sucre blanc/sucre complet (apportant un petit goût de caramel). Enfin, nous avons repéré 3 facteurs liés au déroulement de la recette : l'utilisation d'un fouet électrique ou manuel, le repos ou non avant la cuisson et la température de cuisson.

Partie 1 : Plan fractionnaire

Dans cette partie, on définit (à l'aide des recettes trouvées sur internet) 2 modalités pour chacune des 7 variables. Voici le tableau récapitulatif :

Variable	Modalité 1	Modalité 2
Proportion sucre blanc/sucre complet	100% /0%	75%/25%
Proportion chocolat noir/chocolat lait	100%/0%	80%/20%
Repos avant cuisson	1h de repos	Pas de repos
Type de fouettage	Fouettage à la main	Fouet électrique
Cuisson	5 min à 200°C puis 25 min à 120°C	30min à 120°C
Quantité de beurre salé	40g	30g
Quantité de sucre	30g	20g

1. **Combien d'essais sont nécessaire pour réaliser le plan complet de cette expérience.**

Le plan complet construit avec 7 facteurs à 2 modalités nécessite 2^7 soit 128 essais.

2. **Proposer et construire sur R un plan d'expérience réalisable en 8 essais.**

Le plan réalisable en 8 essais est le plan 2^(7-4).

```
library(FrF2)
plan <- FrF2(nruns=8,nfactors=7,factor.names=list(sucre=c("100/0","75/25"),
chocolat=c("100/0","80/20"),repos=c("oui","non") , fouet=c("main","Ã©lectrique"),
cuisson=c("200puis120","120"),qtbeurre=c("40g","30g"),qtsucre=c("30g","20g")))
summary(plan)
Experimental design of type  FrF2
8  runs

Factor settings (scale ends):

Design generating information:
$legend
[1] A=sucre    B=chocolat C=repos    D=fouet    E=cuisson  F=qtbeurre G=qtsucre

$generators
[1] D=AB  E=AC  F=BC  G=ABC


Alias structure:
$main
[1] A=BD=CE=FG B=AD=CF=EG C=AE=BF=DG D=AB=CG=EF E=AC=BG=DF F=AG=BC=DE G=AF=BE=CD


The design itself:
class=design, type= FrF2
```

	sucre <fctr>	chocolat <fctr>	repos <fctr>	fouet <fctr>	cuisson <fctr>	qtbeurre <fctr>	qtsucre <fctr>
1	75/25	100/0	non	main	120	40g	30g
2	75/25	100/0	oui	main	200puis120	30g	20g
3	75/25	80/20	non	Ã©lectrique	120	30g	20g
4	100/0	80/20	non	main	200puis120	30g	30g
5	100/0	100/0	oui	Ã©lectrique	120	30g	30g
6	100/0	80/20	oui	main	120	40g	20g
7	75/25	80/20	oui	Ã©lectrique	200puis120	40g	30g
8	100/0	100/0	non	Ã©lectrique	200puis120	40g	20g

8 rows

3. **Commenter la qualité de ce plan et les paramètres que l'on pourrait estimer.**

Ce plan est saturé car il y a autant de facteurs à estimer (7 variables + la constante) que de données. Il n'y a donc pas d'estimation de la résiduelle et on ne peut estimer aucune interaction de second ordre et plus entre les effets. Il faut donc réaliser au moins 16 expériences pour étudier ces 7 facteurs.

PARTIE 2 : PLAN OPTIMAL

Dans cette deuxième partie, nous allons essayer de trouver un plan optimal en moins de 15 expériences afin de pouvoir les réaliser en pratique. On trouve également qu'il manque une modalité pour les variables chocolat et sucre. Voici le tableau récapitulatif avec l'ajout des nouvelles modalités.

Variable	**Modalité 1**	**Modalité 2**	**Modalité 3**
Proportion sucre blanc/sucre complet	100% /0%	80%/20%	60%/40%
Proportion chocolat noir/chocolat lait	100%/0%	80%/20%	60%/40%
Repos avant cuisson	1h de repos	Pas de repos	
Type de fouettage	Fouettage à la main	Fouet électrique	
Cuisson	5 min à 200°C puis 25 min à 120°C	30min à 120°C	
Quantité de beurre salé	40g	30g	
Quantité de sucre	30g	20g	

4. Construire le plan complet sur R et donner son nombre d'essais.

```
library(DoE.base)
library(AlgDesign)
plancomplet<-fac.design(nlevels =
c(3,3,2,2,2,2,2),factor.names=list(sucre=c("100/0","80/20","60/40"),
chocolat=c("100/0","80/20","60/40"),repos=c("non","1h"), fouet=c("main","electrique"),
cuisson=c("200puis120","120"),qtbeurre=c("40g","30g"),qtsucre=c("30g","20g")))
```

Le plan complet comporte 288 essais.

5. Combien d'essais doit-on réaliser au minimum et pour avoir l'orthogonalité entre les facteurs principaux ?

Il y a 5 variables à 2 facteurs, 2 variables à 3 facteurs et la constante à estimer ce qui fait 5*1+2*2+1=10 essais à faire au minimum.

Pour avoir l'orthogonalité parfaite entre les facteurs, il faut un nombre d'essais égal à PPCM(2*2,2*3,3*3) = PPCM(4,6,9) soit 36 essais.

6. Construire un plan optimal en 12 essais.

```
set.seed(1234)
optimal<-optFederov(~.,data=plancomplet,nTrials=12)
optimal$design
```

	sucre	chocolat	repos	fouet	cuisson	qtbeurre	qtsucre
	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>
29	100/0	60/40	non	main	200puis120	30g	20g
31	60/40	60/40	1h	electrique	200puis120	30g	30g
77	80/20	60/40	1h	main	120	40g	30g
132	80/20	100/0	non	electrique	200puis120	30g	30g
141	60/40	80/20	non	main	120	30g	30g
161	60/40	60/40	non	electrique	120	40g	20g
248	60/40	100/0	1h	main	200puis120	40g	20g
253	100/0	80/20	1h	electrique	200puis120	40g	30g
265	100/0	100/0	non	electrique	120	40g	30g
274	100/0	100/0	1h	main	120	30g	20g
279	80/20	80/20	non	main	200puis120	40g	20g
288	80/20	80/20	1h	electrique	120	30g	20g

12 rows

7. Construire et commenter la matrice de dispersion de ce plan.

```
X <- model.matrix(~. ,data=optimal$design)
S<-solve(t(X)%*%X)
round(S,2)
```

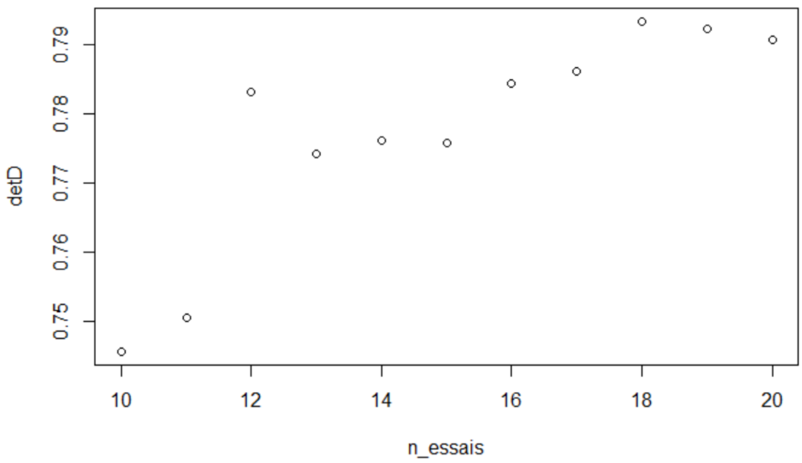
```
            (Intercept) sucre1 sucre2 chocolat1 chocolat2 repos1 fouet1 cuisson1 qtbeurre1 qtsucre1
(Intercept)        0.08   0.00   0.00      0.00      0.00   0.00   0.00     0.00      0.00     0.00
sucre1             0.00   0.18  -0.09     -0.04      0.02   0.00   0.00     0.00      0.00     0.00
sucre2             0.00  -0.09   0.18      0.02     -0.04   0.00   0.00     0.00      0.00     0.00
chocolat1          0.00  -0.04   0.02      0.18     -0.09   0.00   0.00     0.00      0.00     0.00
chocolat2          0.00   0.02  -0.04     -0.09      0.18   0.00   0.00     0.00      0.00     0.00
repos1             0.00   0.00   0.00      0.00      0.00   0.08   0.00     0.00      0.00     0.00
fouet1             0.00   0.00   0.00      0.00      0.00   0.00   0.09     0.00      0.00     0.03
cuisson1           0.00   0.00   0.00      0.00      0.00   0.00   0.00     0.08      0.00     0.00
```

```
qtbeurre1          0.00   0.00   0.00     0.00     0.00   0.00   0.00    0.00     0.08    0.00
qtsucre1           0.00   0.00   0.00     0.00     0.00   0.00   0.03    0.00     0.00    0.09
```

On voit que la matrice de dispersion, ici construite seulement avec les effets principaux, n'est pas parfaitement diagonale. Cela veut dire que les modalités ne sont pas parfaitement orthogonales entre elles, qu'il peut y avoir des confusions. Ce sont surtout les modalités des facteurs à 3 modalités qui sont confondus entre eux.

8. **Réaliser un graphique qui représente la valeur du déterminant D en fonction du nombre n d'essais choisi. On prendra 10 valeurs de n à partir du nombre minimal d'essais déterminé dans la question 2. Pour combien d'essais a-t-on le meilleur déterminant D ?**

```
complet<-fac.design(nlevels = c(3,3,2,2,2,2,2))
detD<-NULL
n_essais<-c(10:20)
for (i in 10:20) {
  set.seed(1234)
  optimal<-optFederov(~.,data=complet,nTrials=i)
  detD <- c(detD,optimal$D)
}
plot(detD~n_essais)
```




```
m<-max(detD)
m
[1] 0.79333

which(detD==m)+9
[1] 18
```

Le plan avec 18 essais est le meilleur avec un déterminant D égal à 0,793.

Par faute de temps et de moyens, on décide de réaliser le plan optimal à 12 essais. On fait déguster les gâteaux à 2 juges de la manière suivante : ils doivent mettre une note sur 15 comprenant 3 notes sur 5 (visuel, texture et goût) visant à comparer chaque gâteau de l'expérience avec le vrai fondant Baulois. L'ordre de dégustation est aléatoire donc différent pour les 2 juges.

9. **Expliquer comment on pourra analyser les réusltats du plan, sachant qu'il faut prendre en compte l'influence des 2 déustateurs différents. Préciser le modèle utilisé.**

Afin de prendre en compte l'influence des 2 juges, il faudra faire une analyse de variance à 8 facteurs : les 7 facteurs étudiés pour l'expérience + le facteur juge à 2 modalités. Le tableau de donnée pour l'analyse de variance contiendra ainsi 24 lignes au lieu de 12 pour le plan d'expérience.

Le modèle étudié s'écrira donc :

$$Y_{ijklmnop} = \mu_i + \alpha_j + \beta_k + \gamma_l + \delta_m + \theta_n + \vartheta_o + \tau_p + \varepsilon_{ijklmnop}$$

Avec $i = 1,2,3$; $j = 1,2,3$; $k = 1,2$; $l = 1,2$; $m = 1,2$; $n = 1,2$; $o = 1,2$; $p = 1,2$

Et $\mathcal{L}(\varepsilon_{ijklmnop}) = \mathcal{N}(0, \sigma)$

Voici les notes obtenues, dans l'ordre pour les 2 juges :

Juge n°1 : 8,6,6,7.5,6,4,7,10.5,4.5,4,9,7

Juge n°2 : 9,8,6,4,8,5,8,7,7,10,10,5

10. Saisir les notes et faire l'analyse sur R en utilisant la fonction `AovSum` puis commenter les résultats.

```
#importation du plan experimental
planexp<-read_excel("planexp_baulois.xlsx")
#duplication
planexp<-planexp[rep(1:nrow(plan),each=2),]
#variable juge
juge <- as.factor(c(1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2))
#variable reponse : notes des juges 1 et 2
Y<-c(8,9,6,8,6,6,7.5,4,6,8,4,5,7,8,10.5,7,4.5,7,4,10,9,10,7,5)
#ajout des variables juge et Y au jeu de donnees
planexp <- cbind.data.frame(planexp,juge,Y)
for (i in 1:9) planexp[,i] <- as.factor(planexp[,i]) #mise en facteur
#construction du modele
AovSum(Y~sucre+chocolat+repos+fouet+cuisson+qtbeurre+qtsucre+juge,data=planexp)
```

```
Ftest
             SS df      MS F value  Pr(>F)
sucre     5.779  2  2.8896  0.9226 0.42201
chocolat 10.779  2  5.3896  1.7208 0.21726
repos     0.260  1  0.2604  0.0831 0.77762
fouet     8.755  1  8.7552  2.7954 0.11842
cuisson  19.260  1 19.2604  6.1496 0.02762 *
qtbeurre  0.094  1  0.0938  0.0299 0.86531
qtsucre   0.005  1  0.0052  0.0017 0.96809
juge      2.344  1  2.3438  0.7483 0.40269
Residuals 40.716 13  3.1320
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ttest
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           6.937500   0.361246 19.2044  < 2e-16 ***
sucre - 100/0         0.716667   0.527634  1.3583  0.19748
sucre - 60/40        -0.350000   0.527634 -0.6633  0.51870
sucre - 80/20        -0.366667   0.527634 -0.6949  0.49933
chocolat - 100/0     -0.616667   0.527634 -1.1687  0.26349
chocolat - 60/40     -0.350000   0.527634 -0.6633  0.51870
chocolat - 80/20      0.966667   0.527634  1.8321  0.08994 .
repos - 1h            0.104167   0.361246  0.2884  0.77762
repos - non          -0.104167   0.361246 -0.2884  0.77762
fouet - electrique   -0.640625   0.383159 -1.6720  0.11842
fouet - main          0.640625   0.383159  1.6720  0.11842
cuisson - 120        -0.895833   0.361246 -2.4798  0.02762 *
cuisson - 200puis120  0.895833   0.361246  2.4798  0.02762 *
qtbeurre - 30g       -0.062500   0.361246 -0.1730  0.86531
qtbeurre - 40g        0.062500   0.361246  0.1730  0.86531
qtsucre - 20g         0.015625   0.383159  0.0408  0.96809
qtsucre - 30g        -0.015625   0.383159 -0.0408  0.96809
juge - 1             -0.312500   0.361246 -0.8651  0.40269
juge - 2              0.312500   0.361246  0.8651  0.40269
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Malheureusement, on voit sur le tableau d'analyse de variance, le test F, que seulement le facteur « cuisson » est significatif. On ne peut donc pas, pour tous les autres facteurs, rejeter l'hypothèse H0 et donc interpréter l'estimation des coefficients. Si on regarde tout de même pour le facteur « cuisson », on voit dans le tableau des coefficients, le test T, que c'est une cuisson en 2 temps (à 200°C puis 120°C) qui augmente la note du gâteau de presque 1 point.

On peut expliquer ce manque de résultats de différentes manières. Il était peut-être compliqué de noter les gâteaux de manière objective pour les juges, sachant que les gâteaux avaient des recettes proches. Il aurait peut-être fallu faire déguster à plus de juges. Enfin, il aurait été préférable de réaliser plus d'essais pour une meilleure orthogonalité.

<u>Annexe 1 :</u> Organisation des expériences

- Sans repos ; 200°C puis 120°C

N°29 6 / 5	N°132 5 / 4	N°279 7 / 6
- 1 œuf - 5g farine - 30g beurre	- 1 œuf - 5g farine - 30g beurre	- 1 œuf - 5g farine - 40g beurre

- 20g sucre blanc - 24g chocolat noir - 16g chocolat lait - Fouet à la main	- 24g sucre blanc - 6g sucre complet - 40g chocolat noir - Fouet électrique	- 16g sucre blanc - 4g sucre complet - 32g chocolat noir - 8g chocolat lait - Fouet à la main

- Avec repos ; 120°C

N°77 7 / 5,5	N°274 5 / 3	N°288 6 / 6
- 1 œuf - 5g farine - 40g beurre - 24g sucre blanc - 6g sucre complet - 24g chocolat noir - 16g chocolat lait - Fouet à la main	- 1 œuf - 5g farine - 30g beurre - 20g sucre blanc - 40g chocolat noir - Fouet à la main	- 1 œuf - 5g farine - 30g beurre - 16g sucre blanc - 4g sucre complet - 32g chocolat noir - 8g chocolat lait - Fouet électrique

- Avec repos ; 200°C puis 120°C

N°31 9 / 8,5	N°248 8 / 7	N°253 7,5 / 6,5
- 1 œuf - 5g farine - 30g beurre - 18g sucre blanc - 12g sucre complet - 24g chocolat noir - 16g chocolat lait - Fouet électrique	- 1 œuf - 5g farine - 40g beurre - 12g sucre blanc - 8g sucre complet - 40g chocolat noir - Fouet à la main	- 1 œuf - 5g farine - 40g beurre - 30g sucre blanc - 32g chocolat noir - 8g chocolat lait - Fouet électrique

- Sans repos ; 120°C

N°141 7,5 / 6	N°161 8,5/8	N°265 5/4
- 1 œuf - 5g farine - 30g beurre - 18g sucre blanc - 12g sucre complet - 32g chocolat noir - 8g chocolat lait - Fouet à la main	- 1 œuf - 5g farine - 40g beurre - 12g sucre blanc - 8g sucre complet - 24g chocolat noir - 16g chocolat lait - Fouet électrique	- 1 œuf - 5g farine - 40g beurre - 30g sucre blanc - 40g chocolat noir - Fouet électrique

<u>Annexe 2</u> : Photo des 12 gateaux réalisés

Exercice 21. Produire des tomates savoureuses (exemple fictif, par M. PLU)

Thématique fictive, pas de données réelles pour cet exercice.

Un producteur martiniquais de tomates hors-sol cherche à maximiser son rendement. Pour cela, il veut réaliser des expériences afin de définir l’itinéraire technique le plus productif. Le rendement d’un plant de tomate peut dépendre de nombreux facteurs comme la variété, l’irrigation, la fertilisation, les traitements et la densité.

Ensuite, il cherche à avoir des tomates savoureuses. Pour cela, il veut réaliser des expériences afin de définir les paramètres optimaux pour récolter des tomates savoureuses. Le gout d’une tomate peut dépendre de facteurs comme la durée d’ensoleillement journalière, l’irrigation et la température moyenne journalière.

Ce producteur ne peut consacrer tout l’espace de sa serre aux expériences car il souhaite également continuer son activité de production avec d’autres cultures. Il dispose d’un espace limité lui permettant de réaliser seulement 16 essais pour l’expérience concernant le rendement et 22 essais pour l’expérience concernant le gout des tomates.

Partie 1 : Quels sont les modalites optimales pour avoir un rendement en tomate maximise ?

Pour répondre à sa problématique, il décide de réaliser un plan d’expérience à partir des facteurs suivants :

Facteurs et modalités à tester (description des modalités en annexe 1) :

- A : Varieté (var1 ; var 2)
- B : Fertilisation (sol1 ; sol2)
- C : Traitement (préventif ; curatif)
- D : Densité (faible ; forte)
- E : Irrigation (homogène ; progressive)

1. **En fonction des facteurs et modalités qu’il souhaite tester, quel est le type de plan à utiliser ? Quelles sont les propriétés de ce plan ?**

   Il doit utiliser un plan fractionnaire. Il s’agit d’un plan 2^{p-k}, 2^{5-1}, soit un plan en 16 essais. (5 facteurs à 2 modalités)

2. **Donner les propriétés de ce plan en respectant les contraintes de l'agriculteurs : quelles sont les confusions du plan ? sa résolution ? qu'est-ce que cela signifie ?**

Contraintes : 16 essais maximum et toutes les interactions d'ordre 2 doivent être estimées.

Confusion réalisée : E = ABCD

I	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD
CDE	BCDE	ACDE	ABDE	ABCE	CDE	BDE	BCE	ADE	ACE	ABE	DE	CE

Les effets principaux du plan sont confondus avec des interactions d'ordre 4 et les effets secondaires avec des interactions d'ordre 3. Il s'agit donc d'un plan avec une résolution d'ordre 5.

Cela signifie que tous les effets principaux et toutes les interactions d'ordre 2 peuvent être estimées.

Ce plan est réalisable en 16 essais.

3. **Quelles sont donc les essais à tester ?**

```
library(FrF2)
plan2 <- FrF2(nruns=16, nfactors=5,
factor.names=list(var=c("var1","var2"),ferti=c("sol1","sol2"),
traitement=c("preventif","curatif"),densite=c("faible","forte"),irrigation=c("homogène","progres
sive")))
summary(plan2)
Call:
FrF2(nruns = 16, nfactors = 5, factor.names = list(var = c("var1",
    "var2"), ferti = c("sol1", "sol2"), traitement = c("preventif",
    "curatif"), densite = c("faible", "forte"),
    irrigation = c("homogène", "progressive")))

Experimental design of type  FrF2
16  runs

Factor settings (scale ends):
   var ferti traitement densite  irrigation
1 var1  sol1  preventif  faible    homogène
2 var2  sol2    curatif   forte progressive

Design generating information:
$legend
[1] A=var        B=ferti      C=traitement D=densite    E=irrigation

$generators
[1] E=ABCD


Alias structure:
[[1]]
[1] no aliasing among main effects and 2fis


The design itself:
    var ferti traitement densite  irrigation
1  var2  sol2  preventif   forte    homogène
2  var2  sol2    curatif   forte progressive
3  var1  sol1  preventif  faible progressive
4  var2  sol1    curatif  faible progressive
5  var1  sol1    curatif  faible    homogène
6  var2  sol1  preventif   forte progressive
7  var2  sol2  preventif  faible progressive
8  var1  sol2  preventif  faible    homogène
9  var2  sol2    curatif  faible    homogène
10 var1  sol1    curatif   forte progressive
11 var2  sol1    curatif   forte    homogène
12 var1  sol2    curatif   forte    homogène
13 var1  sol2  preventif   forte progressive
14 var1  sol2    curatif  faible progressive
15 var1  sol1  preventif   forte    homogène
16 var2  sol1  preventif  faible    homogène
class=design, type= FrF2
```

4. **Quel modèle sera à utiliser pour analyser ce plan**

Pour analyser les résultats, un modèle ANOVA à 5 facteurs sera construit :

$Y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \zeta_m + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\delta)_{il} + (\alpha\zeta)_{im} + (\beta\gamma)_{jk} + (\beta\delta)_{jl} + (\beta\zeta)_{jm} + (\gamma\delta)_{kl} + (\gamma\zeta)_{km} + (\delta\zeta)_{lm} + \varepsilon_{ijklm}$

α_i : l'effet de la variété

β_j : l'effet de la fertilisation

γ_k : l'effet du type de traitement

δ_l : l'effet de la densité de semis

ζ_m : l'effet de l'irrigation

ε_{ijklm} : la résiduelle

PARTIE 2 : QUELS SONT LES PARAMETRES OPTIMAUX POUR RECOLTER DES TOMATES SAVOUREUSES ?

Ce producteur de tomate aimerait que ses tomates soient savoureuses. Son conseiller agricole lui a dit que le gout sucré des tomates dépendrait principalement de 3 facteurs : l'irrigation, la durée d'ensoleillement et la moyenne des températures journalières.

Les diverses plages d'utilisation possibles sont :

- A : Durée d'ensoleillement (12h ; 20h)
- B : Irrigation (10 irrigations/jour ; 16 irrigations/jour)
- C : Moyenne de température journalière (24°C ; 30°C)

Pour répondre à sa problématique, il décide de réaliser un plan composite centré :

5. **Proposer et réaliser un tel plan.**

On construit un plan composite centré. Ici, $\alpha = \sqrt[4]{8}$ et $n\alpha = 6$ et enfin, le nombre de points au centre est $n0 = 8$.

Total d'essais = $2^3+6+8=22$. 22 essais sont à réaliser.

```
library(rsm)
plan <- ccd(3)
plan<-ccd(3, coding=list (x1~(dureeSol-16)/4, x2~(irrigation-13)/3,x3~(temperature-27)/3))
plan
   run.order std.order  dureeSol irrigation temperature Block
1          1        10 16.000000  13.000000    27.00000     1
2          2         2 20.000000  10.000000    24.00000     1
3          3         3 12.000000  16.000000    24.00000     1
4          4         8 20.000000  16.000000    30.00000     1
5          5        11 16.000000  13.000000    27.00000     1
6          6         1 12.000000  10.000000    24.00000     1
7          7         5 12.000000  10.000000    30.00000     1
8          8         9 16.000000  13.000000    27.00000     1
9          9        12 16.000000  13.000000    27.00000     1
10        10         6 20.000000  10.000000    30.00000     1
11        11         4 20.000000  16.000000    24.00000     1
12        12         7 12.000000  16.000000    30.00000     1
13         1         1  8.697033  13.000000    27.00000     2
14         2         9 16.000000  13.000000    27.00000     2
15         3         7 16.000000  13.000000    27.00000     2
16         4         3 16.000000   7.522774    27.00000     2
17         5         5 16.000000  13.000000    21.52277     2
18         6         8 16.000000  13.000000    27.00000     2
19         7         6 16.000000  13.000000    32.47723     2
20         8        10 16.000000  13.000000    27.00000     2
21         9         2 23.302967  13.000000    27.00000     2
22        10         4 16.000000  18.477226    27.00000     2

Data are stored in coded form using these coding formulas ...
x1 ~ (dureeSol - 16)/4
x2 ~ (irrigation - 13)/3
x3 ~ (temperature - 27)/3
```

6. **Décrire à l'expérimentateur 3 expériences qu'il va réaliser (une par partie du PCC : factoriel, étoiles, centre). Préciser l'intérêt de réaliser des expériences avec des données au centre des facteurs.**

Factoriel : 20 heures d'ensoleillement, 10 irrigations par jours et une température de 24°C en moyenne par jour

Étoile : 16 heures d'ensoleillement, 13 irrigations par jours et une température de 21,52°C en moyenne par jour

Centré : 16 heures d'ensoleillement, 13 irrigations par jours et une température de 27°C en moyenne par jour

L'intérêt des points au centre : Les modalités optimales peuvent être les valeurs au centre, le plan peut prendre différentes formes, notamment la forme parabolique avec des optimums pour les valeurs centrées des modalités.

7. **Commentez la qualité du plan.**

```
plan<-ccd(3, coding=list (x1~(dureeSol-16)/4, x2~(irrigation-13)/3,x3~(temperature-27)/3))
X<-model.matrix(~x1+x2+x3+I(x1^2)+I(x2^2)+I(x3^2)+I(x1*x2)+I(x1*x3)+I(x2*x3)+I(x1*x2*x3),data=pl
an)
t(X)%*%X
                 (Intercept)     x1     x2     x3 I(x1^2) I(x2^2) I(x3^2) I(x1 * x2) I(x1 * x3) I(x2 * x3) I(x1 * x2 * x3)
(Intercept)           22.000  0.000  0.000  0.000  14.667  14.667  14.667          0          0          0               0
x1                     0.000 14.667  0.000  0.000   0.000   0.000   0.000          0          0          0               0
x2                     0.000  0.000 14.667  0.000   0.000   0.000   0.000          0          0          0               0
x3                     0.000  0.000  0.000 14.667   0.000   0.000   0.000          0          0          0               0
I(x1^2)               14.667  0.000  0.000  0.000  30.222   8.000   8.000          0          0          0               0
I(x2^2)               14.667  0.000  0.000  0.000   8.000  30.222   8.000          0          0          0               0
I(x3^2)               14.667  0.000  0.000  0.000   8.000   8.000  30.222          0          0          0               0
I(x1 * x2)             0.000  0.000  0.000  0.000   0.000   0.000   0.000          8          0          0               0
I(x1 * x3)             0.000  0.000  0.000  0.000   0.000   0.000   0.000          0          8          0               0
I(x2 * x3)             0.000  0.000  0.000  0.000   0.000   0.000   0.000          0          0          8               0
I(x1 * x2 * x3)        0.000  0.000  0.000  0.000   0.000   0.000   0.000          0          0          0               8
```

La matrice est presque parfaite. En effet, la constante est indépendante des effets principaux et des effets des interactions et légèrement liée aux effets quadratiques. Les effets principaux sont indépendants de tous les autres effets et de la constante. Les effets quadratiques sont légèrement liés aux autres effets quadratiques et les effets des interactions sont indépendants.

Cette matrice va donc nous permettre d'estimer tous les coefficients sans qu'il n'y ait de corrélations entre les coefficients. La qualité du plan est donc bonne.

8. **Les expériences sont réalisées et une note de gout sucré est mise pour une tomate provenant de chaque essai. Les réponses sont collectées dans un vecteurs Y=c(8,10,4,7,2,5,8,5,6,3,6,10,8,3,6,1,4,7,8,10,6,9). A l'aide de la fonction rsm du package rsm, étudier les résultats de votre modèle. Quel modèle conservez-vous ?**

```
library(rsm)
set.seed(1234)
plan<-ccd(3, coding=list (x1~(dureeSol-16)/4, x2~(irrigation-13)/3,x3~(temperature-27)/3))
notes <- c(8,10,4,7,2,5,8,5,6,3,6,10,8,3,6,1,4,7,8,10,6,9)
reg <- rsm(notes~SO(x1,x2,x3),data=plan)
summary(reg)
Call:
rsm(formula = notes ~ SO(x1, x2, x3), data = plan)

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  6.852871   0.812907  8.4301 2.191e-06 ***
x1           1.408056   0.601813  2.3397   0.03741 *
x2           1.200405   0.601813  1.9946   0.06930 .
x3           0.702475   0.601813  1.1673   0.26577
x1:x2       -0.625000   0.814859 -0.7670   0.45790
x1:x3        0.875000   0.814859  1.0738   0.30403
x2:x3        0.625000   0.814859  0.7670   0.45790
x1^2        -0.035526   0.514004 -0.0691   0.94604
x2^2        -0.635526   0.514004 -1.2364   0.23996
x3^2        -0.335526   0.514004 -0.6528   0.52621
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.5551,     Adjusted R-squared:  0.2214
F-statistic: 1.664 on 9 and 12 DF,  p-value: 0.2026

Analysis of Variance Table

Response: notes
                Df Sum Sq Mean Sq F value  Pr(>F)
FO(x1, x2, x3)   3 57.450 19.1501  3.6051 0.04593
TWI(x1, x2, x3)  3 12.375  4.1250  0.7765 0.52929
PQ(x1, x2, x3)   3  9.704  3.2346  0.6089 0.62190
Residuals       12 63.744  5.3120
Lack of fit      5 26.869  5.3737  1.0201 0.47211
Pure error       7 36.875  5.2679

Stationary point of response surface:
        x1         x2         x3
-2.5265164  1.9954522 -0.3890443

Stationary point in original units:
   dureeSol  irrigation temperature
   5.893934   18.986356   25.832867
```

```
Eigenanalysis:
eigen() decomposition
$values
[1]  0.2843456 -0.2112529 -1.0796717

$vectors
          [,1]       [,2]       [,3]
x1  0.8360791 -0.3231024 -0.4433696
x2 -0.1008349  0.7038971 -0.7031082
x3  0.5392625  0.6325612  0.5559337
```

Pour les effets linéaires, seule la durée d'ensoleillement est significative au seuil de 5 % et l'irrigation significative au seuil de 10%. De plus, le test global sur les effets linéaires permet de considérer que les effets linéaires sont significatifs (p-value = 0,04593).

Aucun autre effet n'et significatif (hormis la constante).

On conserve donc un modèle linéaire.

9. Commenter l'erreur d'ajustement

L'erreur d'ajustement du modèle montre qu'on ne peut pas rejeter l'hypothèse que le modèle a une mauvaise qualité d'ajustement car p-value =0,47211.

10. Ajuster le modèle. Conclure sur les paramètres optimaux pour obtenir des tomates savoureuses.

```
reg <- rsm(notes~FO(x1,x2,x3),data=plan)
summary(reg)
```

```
Call:
rsm(formula = notes ~ FO(x1, x2, x3), data = plan)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  6.18182    0.46554 13.2789 9.723e-11 ***
x1           1.40806    0.57016  2.4696   0.02376 *
x2           1.20040    0.57016  2.1054   0.04957 *
x3           0.70248    0.57016  1.2321   0.23378
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.401,     Adjusted R-squared:  0.3011
F-statistic: 4.016 on 3 and 18 DF,  p-value: 0.02372

Analysis of Variance Table

Response: notes
               Df Sum Sq Mean Sq F value  Pr(>F)
FO(x1, x2, x3)  3 57.450 19.1501  4.0164 0.02372
Residuals      18 85.822  4.7679
Lack of fit    11 48.947  4.4498  0.8447 0.61512
Pure error      7 36.875  5.2679

Direction of steepest ascent (at radius 1):
       x1        x2        x3
0.7114421 0.6065231 0.3549365

Corresponding increment in original units:
   dureeSol  irrigation temperature
   2.845768    1.819569    1.064809
```

Une fois le modèle ajusté en tenant compte de l'effet linéaire de x1 (soit la durée d'ensoleillement), x2 (irrigation) et de l'erreur d'ajustement, on peut conclure sur les paramètres optimaux à utiliser. Le modèle est bon car l'erreur d'ajustement n'est pas significative. Les paramètres optimaux sont :

- 2,8 h pour la durée d'ensoleillement
- 1,8 irrigations par jour
- 1,1 degrés pour la température moyenne journalière

Les valeurs propres sont toutes positives, ceci signifie donc que pour avoir des tomates savoureuses selon les personnes ayant réalisées l'analyse sensorielle, il faut au minimum respecter les paramètres ci-dessus.

Ces résultats sont surprenants étant donné que les paramètres sont situés en dessous des fourchettes annoncées. Ceci est surement dû au fait que ce sont des données inventées et donc pas forcément réalistes avec ce qu'il se produirait dans une serre de tomates.

Exercice 22. Parasitisme de *Varroa destructor* dans des ruches d'*Apis Mellifera* en Patagonie. (exemple réel, par E. Soderberg)

Une doctorante cherche à apprécier l'influence de facteurs abiotiques sur l'infestation de ruches d'abeilles mellifère par un parasite : le varroa. Pour cela elle dispose de ruches localisées en Patagonie argentine. Un stagiaire doit lui proposer un dispositif expérimental permettant l'influence de différents facteurs sur l'état de santé des ruches.

Partie 1 – Comment l'environnement et le traitement influent sur le parasitisme. (1/5)

Nos ruches sont situées dans différentes matrices urbaines : en ville ou à proximité d'exploitations où du glyphosate est utilisé. Les abeilles sont donc exposées ou non à des néonicotinoïdes, molécules prouvées comme étant néfastes pour ces dernières. Les ruches ont également été traitées ou non contre le parasite avec un traitement à base d'acide formique. Enfin, la température sur le lieu où se situent chaque ruche est notée. En effet, une température élevée favorise le développement du parasite. On note 2 classes : supérieure et inférieure à 27°C.

Le stagiaire cherche à mettre un place un dispositif expérimental optimal permettant d'apprécier l'influence de ces trois facteurs sur l'état de santé des ruches tout en limitant une utilisation inutile de colonies d'abeilles. Dans un premier temps il propose le plan d'action suivant :

Essai	Environnement	Traitement	Température
1	Exploitation	Oui	>
2	Exploitation	Oui	<
3	Exploitation	Non	>
4	Exploitation	Non	<
5	Ville	Oui	>
6	Ville	Oui	<
7	Ville	Non	>
8	Ville	Non	<

1. **A partir de ce design expérimental, définissez le plan complet et le nombre d'essais à réaliser.**

On décide d'attribuer la valeur +1 pour le facteur environnement (A), modalité « exploitation » et -1 pour la modalité « ville ». De même, on attribue la valeur +1 pour le facteur traitement (B) modalité « oui » et -1 pour la modalité « non ». Idem pour la variable température (C) , en suivant le même modèle.

Cela nous permet de créer le plan complet suivant :

A	B	C
+1	+1	+1
+1	+1	-1
+1	-1	+1
+1	-1	-1
-1	+1	+1
-1	+1	-1
-1	-1	+1
-1	-1	-1

On a construit un plan 2^{3-1} car on a 3 facteurs à 2 modalités. On a donc 4 essais à faire.

2. **Créez la matrice des effets du modèle saturé avec le plan de base en explicitant et expliquant le modèle utilisé. Elle devra contenir les effets principaux et interactions.**

On choisit donc un plan de base à 4 essais :

A	B
1	1
1	-1
-1	1
-1	-1

On peut ensuite faire la matrice des effets du modèle saturé :

Modèle avec interaction : $Y_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$

I	A	B	AB
1	+1	+1	+1
1	+1	-1	-1
1	-1	+1	-1
1	-1	-1	+1

3. **Explicitez le choix des confusions par affectation des effets principaux et déterminez les confusions résultantes. Quelle est la résolution du plan ?**

On confond le facteur C restant, température, avec l'interaction AB d'ordre 2. On peut noter C=AB ce qui génère donc d'autres confusions.

Si C = AB, on peut écrire que CC = ABC et finalement I = ABC car si on multiplie un facteur par lui-même cela donne une colonne de 1 ce qui correspond à la colonne I.

On va plus loin en multipliant par A de chaque côté ce qui donne A = BC, A est confondu avec BC

De même, B = AC, B est confondu avec AC et C = AB, C confondu avec AB.

Avec le générateur d'alias : I = ABC on a la matrice suivante :

I	A	B	C	AB	AC	BC	ABC
1	+1	+1	+1	1	1	1	1
1	+1	-1	-1	-1	-1	1	1
1	-1	+1	-1	-1	1	-1	1
1	-1	-1	+1	1	-1	-1	1

L'interaction est d'ordre 3 car une seule interaction est confondue avec I, celle ABC on a donc un plan de résolution III.

```
Library (FrF2)
plan <- FrF2(nfactors=3, resolution=3)
summary(plan)
```

```
Call:
FrF2(nfactors = 3, resolution = 3)

Experimental design of type  FrF2
4  runs

Factor settings (scale ends):
   A  B  C
1 -1 -1 -1
2  1  1  1

Design generating information:
$legend
[1] A=A B=B C=C

$generators
[1] C=AB


Alias structure:
$main
[1] A=BC B=AC C=AB


The design itself:
```

```
   A  B  C
1 -1 -1  1
2 -1  1 -1
3  1  1  1
4  1 -1 -1
class=design, type= FrF2
```

PARTIE 2 – PRENONS EN COMPTE LA LOCALISATION ET LA SELECTION DES REINES : AJOUT DE 2 NOUVEAUX FACTEURS ET PASSAGE A UN PLAN ASYMETRIQUE (3/5)

Après réflexion, la chercheuse décide qu'il est probant pour son étude d'étudier deux autres facteurs supplémentaires : la ville dans laquelle sont situées les ruches et la reine sélectionnée. Les colonies peuvent être situées dans 3 villes différentes : Lago Puelo, El Hoyo, El Maiten qui sont situées sur un gradient topographique décroissant. Quant aux reines sélectionnées on a là 4 modalités différentes avec 4 types de reines choisies.

On a donc dès lors 3 facteurs à 2 modalités (Environnement, traitement et température), un facteur à 3 modalités : ville et un facteur à 4 modalités : reines sélectionnées

4. **Estimez le nombre d'essais à réaliser et l'orthogonalité des facteurs**

Nombre d'essais max : 2^3 x 3 x 4 = 96

Nombre d'essais min : 1 + 3 x (2 – 1) + (3 – 1) + (4 – 1) = 9

Il faut donc réaliser entre 96 et 9 essais.

Pour l'orthogonalité des facteurs il faut que toutes les combinaisons soient testées un même nombre de fois :

F1 orthogonal à F2, F1 orthogonal à F3 et F2 orthogonal à F3 donc minimum 2 x 2 = 4 essais

F1 orthogonal à F4, F2 orthogonal à F4 et F3 orthogonal à F4 donc minimum 2 x 3 = 6 essais

F1 orthogonal à F5, F2 orthogonal à F5 et F3 orthogonal à F5 donc 2 x 4 = 8 essais

F4 orthogonal à F5 donc 3 x 4 = 12 essais

Il faut trouver le plus petit commun multiple de ces 4 nombres : 24 < 96

Il serait donc possible de demander au logiciel d'utiliser des plans orthogonaux.

5. **Définissez le nom du plan La2b3C avec a le nombre d'essais pour estimer le modèle à 5 facteurs et ce sans confusion d'effets soit sans orthogonalité.**

On peut rendre le facteur à 4 modalités en 2 facteurs à 2 modalités. On a donc 5 facteurs à 2 modalités et 1 facteur à 3 modalités.

SI on s'intéresse aux nombres d'essais : 1 + 5 x (2-1) + 1 x (3 – 1)= 8 essais minimum et pour le nombre d'essais maximum : 25 x 3 = 96.

Le PPCM(8,96) = 8 pour avoir orthogonalité. Pour avoir une planification expérimentale avec 96 essais il faut créer un plan fractionnaire avec nos 5 facteurs à 2 modalités et le plan complet 31 on a donc que L82531.

6. **Déterminez le plan optimal La2b3C à l'aide de R entre les deux propositions réalisées en question 1 et 2.**

La chercheuse veut limiter le nombre d'expérience à réaliser et veut également un plan expérimental optimum.

On a vu qu'on pouvait réaliser cela avec 24 ou 8 essais. Pour voir quel plan est de meilleure qualité on va créer la matrice $(X'X)^{-1}$ diagonale qui permet de limiter la confusion des effets. Ensuite il est possible de comparer le déterminant de la matrice pour 24 et 8 essais.

```
library(FactoMineR)
library(FrF2)
library(DoE.base)
Design.1 <- fac.design(nlevels=c(2,2,2,2,2,3), factor.names= c("A","B","C","D", "E", "F"))
X <- model.matrix(~ . , Design.1)
VARCOV <- solve(t(X)%*%X)
 (Intercept) A1 B1 C1 D1 E1           F.L        F.Q
1             1 -1  1 -1 -1 -1 -7.850462e-17 -0.8164966
2             1  1 -1 -1 -1  1 -7.850462e-17 -0.8164966
3             1  1 -1  1 -1 -1  7.071068e-01  0.4082483
4             1  1  1  1  1  1 -7.071068e-01  0.4082483
5             1  1  1 -1  1  1 -7.071068e-01  0.4082483
6             1  1 -1  1 -1  1  7.071068e-01  0.4082483
7             1 -1  1  1  1 -1 -7.850462e-17 -0.8164966
8             1 -1  1  1  1  1 -7.850462e-17 -0.8164966
```

```
9              1 -1  1 -1  1 -1 -7.071068e-01  0.4082483
[ . . . ]
91             1  1  1  1 -1  1  7.071068e-01  0.4082483
92             1 -1 -1  1  1 -1 -7.850462e-17 -0.8164966
93             1 -1 -1 -1  1 -1 -7.071068e-01  0.4082483
94             1 -1 -1  1  1 -1  7.071068e-01  0.4082483
95             1 -1  1  1 -1  1  7.071068e-01  0.4082483
96             1 -1  1 -1  1  1 -7.850462e-17 -0.8164966
determinant <- det(VARCOV)
show(determinant)
 [1] 1.247592e-15

library(FactoMineR)
library(FrF2)
library (DoE.base)
Design.1.Dopt <- optFederov(~A+B+C+D+E+F,Design.1,nTrials=8,criterion="D")
Xopti <- model.matrix(~ A + B+ C + D + E + F, Design.1.Dopt$design)
show(Xopti)
(Intercept) A1 B1 C1 D1 E1           F.L        F.Q
18             1  1  1 -1  1 -1 -7.850462e-17 -0.8164966
23             1  1 -1  1 -1  1 -7.850462e-17 -0.8164966
25             1 -1  1  1  1  1  7.071068e-01  0.4082483
33             1 -1 -1 -1  1  1 -7.071068e-01  0.4082483
44             1 -1 -1 -1 -1 -1  7.071068e-01  0.4082483
51             1  1 -1  1  1 -1  7.071068e-01  0.4082483
57             1 -1  1  1 -1 -1 -7.071068e-01  0.4082483
92             1 -1 -1  1  1 -1 -7.850462e-17 -0.8164966
det(solve(t(Xopti)%*%Xopti))
[1] 1.271566e-06
```

On obtient les déterminants des matrices et en les comparant on peut déduire le nombre d'essais à réaliser pour avoir un plan expérimental optimal. En effet, cela est celui correspondant à la matrice (X'X)-1 avec le déterminant le plus petit soit le plan complet avec det = 1.247592e-15 contre det = 1.271566e-06 pour le plan optimal en 8 essais.

Exercice 23. Etude de facteurs pouvant influencer les notes d'examen (exemple fictif, par A. Ledoux)

PARTIE 1 : PLAN FRACTIONNAIRE

Lors d'un examen, la note obtenue par un élève dépend de plusieurs facteurs tels que le niveau d'apprentissage, la sévérité du professeur, la difficulté de l'examen, etc. On cherche ici à étudier quelques facteurs qui influencent les notes obtenues par quelques élèves à un examen réalisé un matin.

On s'intéresse à 5 critères différents :

- Le travail : l'élève travaille peu ou il travaille beaucoup
- Le sommeil : l'élève a eu un sommeil court ou un sommeil long
- Le petit déjeuner : l'élève a mangé le matin de l'examen ou non
- L'attitude : en cours, l'élève est généralement attentif ou dissipé
- L'ordre de correction : au début ou à la fin

Pour réaliser cette étude, on décide de faire 8 essais en combinant aléatoirement les critères.

1. **Donner le nom du plan à construire.**

On a 5 facteurs à 2 modalités et on réalise 8 essais. Donc le plan à construire est un plan 2^{5-2}.

2. **Construire le plan sur R. Donner les confusions, les générateurs d'alias et la résolution du plan.**

On rentre les lignes de code suivantes :

```
library(FrF2)
plan1 <- FrF2 (nruns = 8, nfactors = 5, factor.names = list (travail = c("peu", "beaucoup"),
sommeil = c("long", "court"), petit.dejeuner = c("oui", "non"),attitude = c("attentif",
"dissipé"), ordre.correction = c("début", "fin")))
summary(plan1)
```

On obtient les résultats suivants, avec A = travail, B = sommeil, C = petit déjeuner, D = attitude, E = ordre de correction.

Les confusions se trouvent dans `Alias structure` et sont : A=BD=CE, B=AD, C=AE, D=AB, E=AC, BC=DE, BE=CD.

Les générateurs d'alias se trouvent dans `$generators` et sont : D=AB et E=AC, donc I=ABD=ACE.

Les effets principaux sont confondus avec des interactions d'ordre 2, donc c'est un plan de résolution III.

3. **On observe des interactions entre certains facteurs. Avec ce plan, peut-on analyser toutes les interactions ? Si non, combien peut-on en analyser et comment faire pour toutes les analyser ?**

On a construit un plan 2^{5-2}. On peut donc estimer 2^{5-2} = 8 paramètres. On estime ici les 5 facteurs et la constante. Il reste alors 2 paramètres à estimer, on peut estimer 2 interactions.

Si on voulait analyser toutes les interactions, il faudrait augmenter le nombre d'essais.

PARTIE 2 : PLAN POUR SURFACES DE REPONSES

On s'intéresse maintenant plus précisément à trois facteurs :

- Nombre de jours de travail : 5-7
- Durée de travail par jour (en heure) : 2-4
- Durée de sommeil moyenne par nuit (en heure) : 6-8

On aimerait déterminer le ou lesquels de ces facteurs ont un effet sur les notes obtenues. On considère que les élèves sont initialement d'un niveau similaire.

Pour cette étude, on décide de construire un PCC en réalisant autant d'essais que possible.

4. **Proposer un plan adapté à l'étude en précisant le nombre d'essais réalisés.**

On réalise un PCC avec α=1.682, n_α=6 points axiaux et n_0=9 points au centre. On peut réaliser 2^3+2*3+9=23 essais.

5. **On réalise les essais et on créé un vecteur Y contenant les notes obtenues par les élèves. On lance sur R l'analyse du PCC avec les lignes de code suivantes. Commenter les lignes de réponse de R.**

```
library(rsm)
plan2 <- ccd (3, n0=9, randomize=FALSE, coding = list (x1~(jour-6)/2, x2~(trav-3)/2, x3~(som-7)/2))
Y = c (18, 16, 11, 11, 15, 20, 10, 13, 14, 19, 12, 16, 16, 17, 13, 8, 11, 12, 16, 18, 14, 20, 9, 11, 16, 15, 13, 16, 18, 10, 14, 19)
CR.rsm <- rsm(Y~SO(x1, x2, x3), data=plan2)
summary(CR.rsm)
```

```
Call:
rsm(formula = Y ~ SO(x1, x2, x3), data = plan2)

               Estimate  Std. Error t value  Pr(>|t|)
(Intercept)  1.4316e+01  7.4571e-01 19.1981 3.128e-15 ***
x1           8.9746e-01  8.1736e-01  1.0980   0.28408
x2          -2.0928e+00  8.1736e-01 -2.5604   0.01784 *
x3          -1.2395e+00  8.1736e-01 -1.5165   0.14364
x1:x2       -4.7042e-15  1.1214e+00  0.0000   1.00000
x1:x3        1.2500e+00  1.1214e+00  1.1147   0.27702
x2:x3       -3.8244e-15  1.1214e+00  0.0000   1.00000
x1^2        -1.7230e-01  6.2488e-01 -0.2757   0.78533
x2^2         3.9437e-01  6.2488e-01  0.6311   0.53447
x3^2        -3.0631e-02  6.2488e-01 -0.0490   0.96135
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.3485,     Adjusted R-squared:  0.08196
F-statistic: 1.308 on 9 and 22 DF,  p-value: 0.2884

Analysis of Variance Table

Response: Y
                Df  Sum Sq Mean Sq F value  Pr(>F)
FO(x1, x2, x3)   3 101.218  33.739  3.3537 0.03733
TWI(x1, x2, x3)  3  12.500   4.167  0.4142 0.74450
PQ(x1, x2, x3)   3   4.671   1.557  0.1548 0.92549
Residuals       22 221.330  10.060
Lack of fit      5  55.330  11.066  1.1333 0.38062
Pure error      17 166.000   9.765

Stationary point of response surface:
        x1         x2         x3
 0.9695112  2.6533152 -0.4506973

Stationary point in original units:
    jour     trav      som
7.939022 8.306630 6.098605
```

```
Eigenanalysis:
eigen() decomposition
$values
[1]  0.5275371  0.3943694 -0.7304650
```

On s'intéresse d'abord à la table d'analyse de la variance. On remarque que les effets d'interaction et les effets quadratiques ne sont pas significatifs car ils ont une p-value > 5%. Cependant, les effets linéaires ont une p-value < 5%, ce qui signifie qu'au moins un effet linéaire est significatif. L'erreur pure est de 9.765 et la probabilité critique du test sur l'erreur d'ajustement est de 0.38062 > 5% donc le modèle est bien adapté.

6. Ecrire le modèle retenu. Donner la nature de l'optimum et les valeurs de x1, x2 et x3 pour lesquelles il est atteint.

On regarde maintenant en détail l'estimation des coefficients. On remarque que seul l'effet de x2 possède une p-value < 5%, donc c'est le seul effet significatif.

Le modèle retenu est donc : $Y_i = \beta_0 + \beta_2 x_{i2} + \varepsilon_i$

Il y a trois valeurs propres, deux sont positives et une est négative, donc le point stationnaire est un point selle.

L'optimum de Y est obtenu pour x1=7.9, x2=8.3 et x3=6.1.

Annexes : Présentation des modalités des facteurs de la partie 1

Variété :

- Var 1 : Recento
- Var 2 : Tropic boy

Fertilisation :

- Solution nutri 1 =
  Composer la solution mère dans 2 bacs de 70 L.
  rA = 0,5 % rB=0,5 %
  Bac A :
    i. Phosphate monopotassique : 2,5 kg
    ii. Sulfate de Magnésium : 6,0 kg
    iii. Nitrate de Calcium : 13,2 kg
    iv. Sulfate de Potassium : 2,9 kg
  Bac B :
    - Nitrate de Potassium : 1,9 kg

- Kanieltra 10 Fe : 1,5 L

- Solution nutri 2 =
  Composer la solution mère dans 2 bacs de 70 L.
  rA = 0,6 % rB=0,4 %
  Bac A :
  v. Phosphate monopotassique : 2,5 kg
  vi. Sulfate de Magnésium : 6,0 kg
  vii. Nitrate de Calcium : 13,2 kg
  viii. Sulfate de Potassium : 2,9 kg
  Bac B :
  - Nitrate de Potassium : 1,9 kg

Kanieltra 10 Fe : 1,5 L

Densité de semis :

- Faible : 6 plants distants de 40cm
- Forte : 4 plants distants de 60 cm

Irrigation :

- Mode 1 =
  - 10 irrigations de 3 minutes / jour, de la plantation à début floraison
  - 13 irrigations de 3 minutes / jour, de la floraison à récolte
  - 16 irrigations de 3 minutes / jour, pendant la récolte
- Mode 2 =
  - 13 irrigations de 3 minutes / jour, de la plantation à début floraison
  - 13 irrigations de 3 minutes / jour, de la floraison à récolte
  - 13 irrigations de 3 minutes / jour, pendant la récolte